

USSN 09/996,484
G8-US1
8325-2008

I hereby certify that this correspondence is being deposited with the United States Patent and Trademark Office via EFS on April 1, 2009.

4/1/09
Date

Michelle H. H. H.
Signature

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re Application of:

CHOO et al.

Serial No.: 09/996,484

Filing Date: November 28, 2001

Title: MOLECULAR SWITCHES

Examiner: J. Woitach

Group Art Unit: 1636

Confirmation No.: 2713

Customer No.: 20855

REVISED BRIEF ON APPEAL UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is filed pursuant to 37 C.F.R. § 41.37 (see, Fed. Reg. vol. 73. no. 238, page 74972 published December 10, 2008) and is in response to the Final Office Action mailed on June 4, 2008. A Notice of Appeal was received on December 4, 2008, making an Appeal Brief initially due on or before February 4, 2009. A Notice of Non-Compliant Appeal Brief was mailed on March 17, 2009, making a response due on or before April 17, 2009. Accordingly, this Revised Brief is timely filed.

REAL PARTY IN INTEREST

Gendaq Ltd. is the assignee of the instant application, as recorded on August 22, 2005 in the USPTO at Reel 016655, Frame 0867. See, also, Certificate Under 37 C.F.R. § 3.73(b) filed on April 1, 2002. Gendaq, Ltd. is a wholly owned subsidiary of Sangamo BioSciences, Inc. Therefore, the real party in interest is Sangamo BioSciences, Inc.

RELATED APPEALS AND INTERFERENCES

Appellants are not aware of any related appeals or interferences.

STATUS OF CLAIMS

Pending: Claims 1, 2, 4, 5, 7, 8, 10, 11, 13-15, 21-26, 31, 34, 35 and 38-48

Canceled: Claims 3, 6, 9, 12, 16-20, 27-30, 32, 33, 36, 37, 49

Withdrawn: Claims 1, 2, 4, 5, 7, 8, 10, 11, 13-15, 21-26, 31, 35 and 38-47

Rejected: Claims 34 and 48

Appealed: Claims 34 and 48

STATUS OF AMENDMENTS

No amendments have been made subsequent to the mailing of the Final Office Action on June 4, 2008.

Appellants note that their Response after Final was mailed within 2 months of the mailing of the Final Office Action and, therefore, expedited procedure was in order. However, no Advisory Action was ever received, despite repeated telephone calls and a written status inquiry to the Office.

SUMMARY OF CLAIMED SUBJECT MATTER

Independent claim 34 is drawn to a complex (page 10, lines 16-19) comprising (a) a heterodimer comprising first and second polypeptides (page 2, lines 8-11) and (b) a ligand (page 10, lines 18-19). The ligand binds to the first and second polypeptides and mediates heterodimerization of these two polypeptides (page 49, line 25; page 58, lines 12-14; page

59, lines 4-5; paragraph bridging pages 54-55). The first and second polypeptides bind to DNA, and, in addition, the first or second polypeptide comprises an engineered, non-naturally occurring Cys2-His2 zinc finger binding domain (page 23, line 4 through page 31, line 31).

Independent claim 48 is drawn to a switching system comprising a protein switch (page 5, lines 14-15) comprising: (i) a first component comprising a first polypeptide and (ii) a second component comprising a second polypeptide (page 5, lines 15-16), in which the first polypeptide binds to the second polypeptide and the binding of the polypeptides is mediated by a ligand and that binds to both polypeptides (page 5, line 14), and (iii) a third component comprising the ligand, wherein the first and second polypeptides bind to DNA (page 5, lines 18-20), and further wherein the first or second polypeptide comprises an engineered, non-naturally occurring Cys2-His2 zinc finger binding domain (page 23, line 4 through page 31, line 31).

GROUND OF REJECTION TO BE REVIEWED ON APPEAL

A. Whether claims 34 and 48 are unpatentable under 35 U.S.C. § 112, 1st paragraph as not adequately described by the as-filed specification.

B. Whether claims 34 and 48 are unpatentable under 35 U.S.C. § 112, 2nd paragraph as allegedly indefinite.

C. Whether claims 34 and 48 are unpatentable under 35 U.S.C. § 103(a) as obvious in view of WO 96/06110 (hereinafter "Gilman").

ARGUMENTS

A. Claims 34 and 48 are fully described by the as-filed specification

Claims 34 and 48 were rejected under 35 U.S.C. § 112, 1st paragraph as allegedly failing to comply with the written description requirement by containing subject matter that was not described in the originally-filed specification. (Final Office Action, pages 3-4). In particular, it was alleged that the recitation “non-naturally occurring” was not adequately described because naturally occurring DNA binding domains may mutate. *Id.*

For the reasons of record, Appellants reiterate that that the term “non-naturally occurring” is amply described in the as-filed specification. It is well settled that the written description requirement is satisfied if the specification reasonably conveys possession of the invention to one skilled in the art. *See, e.g., In re Lukach*, 169 USPQ 795, 796 (CCPA 1971). The disclosure must be read in light of the knowledge possessed by the skilled artisan at the time of filing, for example as established by reference to patents and publications available to the public prior to the filing date of the application. *See, e.g., In re Lange*, 209 USPQ 288 (CCPA 1981). Moreover, the burden is on the Examiner to provide evidence as to why a skilled artisan would not have recognized that the applicant was in possession of claimed invention at the time of filing. *Vas-Cath, Inc. v. Mahurkar*, 19 USPQ2d 1111 (Fed. Cir. 1991); *In re Wertheim*, 191 USPQ 90 (CCPA 1976).

In the case on appeal, the rejection is premised on the assertion that non-naturally occurring is not adequately described because naturally occurring DNA binding domains can “mutate” spontaneously. (Final Office Action, pages 3-4). However, no evidence is given by the Examiner in support of this assertion, and spontaneous mutations in the binding region of these proteins has not been documented. The line of reasoning followed by the Examiner is completely nonsensical. Following this thinking, patents should never be granted for any novel protein or chemical structure, because there may be a chance that somewhere, in some unknown organism, the novel compound may already have been made, “as a consequence of continual mutation.”

Moreover, the basis of the rejection, namely that naturally occurring zinc finger proteins may mutate spontaneously, is utterly irrelevant to a written description inquiry – an

applicant is not required to present a list of all non-naturally occurring (or all naturally occurring) Cys2-His2 zinc finger proteins in order to satisfy the written description requirement of claims directed to "non-naturally occurring" Cys2-His2 zinc finger proteins. Rather, what is required is that an applicant demonstrates possession of the claimed subject matter. Here, the as-filed specification contains ample description of both naturally and non-naturally occurring Cys2-His2 zinc finger DNA binding domains (see, e.g., paragraphs [0107] and [0120], emphasis added):

A zinc finger binding motif is a structure well known to those in the art and defined in, for example. Miller et al., (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA) 85:99-102; Lee et al. (1989) Science 245:635-637; see International patent applications WO 96/06166 and WO 96/32475, corresponding to U.S. Ser. No. 08/422.107, incorporated herein by reference.

In general, naturally occurring zinc fingers may be selected from those fingers for which the DNA binding specificity is known. For example, these may be the fingers for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson et al., (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall et al., (1993) Nature 366:483-487) and YY1 (Houbaviv et al., (1996) PNAS (USA) 93:13577-13582).

The as-filed specification also clearly describes that non-naturally occurring zinc finger proteins as claimed were known to be obtainable by design or selection at the time of filing (paragraphs [0009], [0039], [0119], [0122], and [0125], emphasis added):

Preferably, at least one of the candidate first molecules comprises a non-naturally occurring binding domain which binds to the second molecule. The term "a non-naturally occurring binding domain" means that the binding domain does not occur in nature, even as part of a larger molecule, and has been obtained by deliberate mutagenesis procedures or de novo design techniques. 94:5525-5530; and Beerli et al. (1998) Proc. Natl. Acad. Sci. USA 95:14628-14633.

As used herein the terms "peptide", "polypeptide" and "protein" refer to a polymer in which the monomers are amino acids and are joined together through peptide or disulfide bonds. "Polypeptide" refers to either a full-length

naturally-occurring amino acid chain or a "fragment thereof" or "peptide", such as a selected region of the polypeptide that binds to another protein, peptide or polypeptide in a manner modulatable by a ligand, or to an amino acid polymer, or a fragment or peptide thereof, **which is partially or wholly non-natural.**

We also describe a method for preparing a DNA binding protein of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence in a manner modulatable by a ligand, comprising the steps of: (a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers: and (b) mutating at least one of positions -1, +3, +6 (and ++2) of the finger as required by a method according to the present invention.

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

When the nucleic acid specificity of the model finger selected is known, the mutation of the finger in order to modify its specificity to bind to the target DNA may be directed to residues known to affect binding to bases at which the natural and desired targets differ. Otherwise, mutation of the model fingers should be concentrated upon residues -1, +3, +6 and ++2 as provided for in the foregoing rules.

Furthermore, the specification need not describe, and preferably omits, that which is well known to the skilled artisan. At the time of filing, the skilled artisan was well that the term non-naturally occurring clearly refers to those Cys2-His2 zinc finger DNA binding domains that had been engineered (e.g., by design, selection or mutagenesis) to bind to a selected target site. See, e.g., Refs. B6, B13 and B16 of the IDS mailed on October 24, 2003 and cited in regards to engineering of non-naturally occurring Cys2-His2 zinc finger binding domains on page 24, lines 18-24 of the as-filed specification.

Thus, the skilled artisan would have no doubt that the as-filed specification, in light of the state of the art at the time of filing, describes non-naturally occurring zinc finger proteins as those that do not occur in nature. Furthermore, the Office has not identified any Cys2-His2 zinc finger DNA binding domains having random, naturally occurring mutations and,

even if such proteins exist, they are not encompassed by the claims because they would be naturally occurring.

Appellants also note that the Board of Patent Appeals and Interferences has recently reaffirmed that the term “naturally occurring” would be understood by the persons of skill in the art to mean that it exists or is found in nature. *See*, page 3 of *Ex parte Dewis et al.* (2007) Appeal 2007-1610 (BPAI), attached hereto. Plainly, the skilled artisan would know that “non-naturally occurring” refers to zinc finger proteins that do not exist or are found in nature.

Since it is clear that the skilled artisan would have known that Appellants were in possession of non-naturally occurring Cys2-His2 zinc finger proteins as claimed, namely by engineering via design or selection to produce a zinc finger protein that does not occur in nature, withdrawal of the rejection is in order.

B. Claims 34 and 48 are clear and definite

Claims 34 and 48 were also rejected under 35 U.S.C. § 112, 2nd paragraph as allegedly indefinite for reciting a “non-naturally occurring Cys2-His2 zinc finger binding domain.” (Final Office Action, pages 4-5).

As detailed in their Response After Final, Appellants traversed the rejection, noting that the term “non-naturally occurring Cys2-His2 zinc finger binding domain” is completely clear to the skilled artisan.

The definiteness requirement of 35 U.S.C. § 112, second paragraph is satisfied if it is clear to the skilled artisan what is meant by a particular claim term. *See, e.g., In re Marosi*, 218 USPQ 289 (Fed. Cir. 1983). The definiteness and clarity of claim language must be analyzed, not in a vacuum, but in light of (1) the content of the particular disclosure; (2) the teachings of the art; and (3) the claim interpretation that would be given by one possessing ordinary skill in the pertinent art at the time the invention was made. *See, e.g., W.L. Gore & Assocs., Inc. v. Garlock, Inc.*, 220 USPQ 202 (Fed. Cir. 1983).

In the case on appeal, the as-filed specification more than clearly defines what is encompassed by the recitation “non-naturally occurring” Cys2-His2 zinc finger protein.

Specifically, as discussed in the record and the instant Brief, the term “non-naturally occurring” clearly refers to any binding domain that does not occur in nature, namely zinc finger proteins which have been altered in the recognition region helix by design or selection to bind to a selected target site. See, also citations from the specification above in Section A regarding 35 U.S.C. § 112, 1st paragraph.

Thus, it is clear from the as-filed specification that the term “non-naturally occurring” refers to a zinc finger protein in which the DNA recognition regions of one or more of the component fingers have been designed or selected for binding to a particular target site.

Finally, the Examiner’s assertion that it is “impossible” to know whether any sequence is naturally occurring because not all naturally occurring proteins are known and because proteins change over time is incorrect and does not support the contention that the claims are indefinite. Zinc finger proteins can be naturally-occurring or they can be non-naturally-occurring; the claims make explicit that the claimed zinc finger DNA-binding domain is non-naturally-occurring. Furthermore, at any point in time, through ordinary searching of the extensive databases now available publically to the artisan, it is a simple and straightforward matter for one of skill in the art to determine what is or is not naturally-occurring; thereby determining what is encompassed by the claims.

Thus, in view of the specification as a whole and state of the art, the claims are clear and withdrawal of the rejection is in order.

C. Claims 34 and 48 are non-obvious over Gilman

Claims 34 and 48 were rejected as allegedly obvious over WO 96/06110 (hereinafter “Gilman”). (Final Office Action, pages 5-9). Gilman was cited for allegedly teaching all the claimed elements except for a non-naturally occurring Cys2-His2 zinc finger binding domain, although “non-naturally occurring” was alleged to be “impossible” to determine and/or encompassed by Gilman’s disclosure of phage display libraries. *Id.*

Again, the rejection cannot be sustained if the term “non-naturally occurring” as applied to Cys2-His2 zinc finger domain is properly interpreted in the context of the claim. For the reasons detailed above, it is entirely clear and definite what is encompassed by the

recitation “non-naturally occurring.” The specification in fact clearly defines what is meant by the claim term.

Moreover, as set forth in *Philips v. AWH*, 75 USPQ2d, 1321, 1326 (Fed. Cir. 2005) (and a host of prior case law¹) the primary determinant of the meaning of a claim term is the ordinary and customary meaning of that term:

the ordinary and customary meaning of a claim term is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention.

The ordinary and customary meaning of the term “non-naturally occurring” is something that does not occur naturally, for example Cys2-His2 zinc finger proteins whose recognition domains had been designed and/or selected (engineered) to bind to a target site of choice. As noted above, evidence has also been provided establishing that the Board considers the ordinary and customary meaning of the term “non-naturally occurring” to be any composition that does not occur in nature. See, *Ex parte Dewis*, Evidence Appendix (1). Further, nothing in the specification contradicts what one of ordinary skill in the art of Cys2-His2 zinc fingers, as of Appellants’ filing date, would consider to be the ordinary and customary meaning of the term “non-naturally occurring.”

The pending claims require that the DNA binding domain be non-naturally occurring and, thus, every naturally occurring DNA binding domain sequence is excluded from the scope of the claims. Thus, as acknowledged Gilman fails to teach or suggest anything about engineered zinc finger proteins in addition to failing to teaching anything about non-naturally occurring Cys2-His2 zinc finger binding domains.

Importantly, Gilman also fails to teach, suggest or enable complexes as claimed in which heterodimerization of first and second DNA binding domains is mediated by a ligand that binds to the DNA binding domains. Rather, Gilman teaches that DNA binding domains

¹ See, e.g., *Vitronics Corp. v. Conception, Inc.*, 90 F.3d 1576 (Fed. Cir. 1996); *Ferguson Beauregard/Logic Controls v. Mega Sys., LLC*, 350 F.3d 1327, 1338 (Fed. Cir. 2003); *Innova Pure Water, Inc. v. Safari Water Filtration Systems, Inc.*, 381 F.3d 1111, 1116 (Fed. Cir. 2004) and *Home Diagnostics, Inc. v. LifeScan, Inc.*, 381 F.3d 1352, 1358 (Fed. Cir. 2004)

are either covalently linked (i.e., via a linker in a fusion protein) (Gilman, page 9) or that fusion proteins containing both a DNA-binding domain and immunophilin ligand-binding domains are linked by a linker that binds to the fused immunophilin domain. Specifically, Gilman discloses that two or more DNA-binding domains are covalently linked via traditional linkers to form a fusion protein. See, Gilman, sections 3 and 4, beginning on page 6 of the disclosure, particularly page 9. Indeed, as previously noted, Gilman clearly links his DNA-binding domains covalently to form “chimeric” or “composite” DNA binding domains. Once covalently linked, a ligand-binding domain may be added an “additional domain” to link two or more composite molecules (page 7, lines 29-36; page 10, lines 17-21; and page 10, lines 22, emphasis added):

The chimeric proteins may also include a ligand-binding domain to provide for regulatable interaction of the protein with a second polypeptide chain. Thus, in embodiments involving covalently linked composite DNA binding domains, the unitary composite DNA-binding protein may further contain a ligand-binding domain. In such cases, the presence of a ligand-binding domain permits association of the composite DBP, in the presence of a dimerizing ligand, with a second chimeric protein containing a transcriptional activation domain and another ligand-binding domain.

Additional domains, described in the previous section (e.g., activation domains, ligand-binding domains) may be appended to either the N- or C-termini of the DNA-binding domains in any order consistent with the proper functioning of the protein

Gilman also only exemplifies complexes in which two DNA-binding domains are covalently linked as a fusion protein. See, Examples of Gilman. This is entirely unlike the claimed complexes in which a ligand modulates formation of a heterodimer.

Moreover, in terms of ligand-mediated multimerization, Gilman also teaches only that this is accomplished by fusing an immunophilin ligand-binding domain to the DNA-binding domain (page 11, lines 1-22 of Gilman, emphasis added):

In embodiments involving composite DNA-binding proteins formed by ligand-mediated multimerization rather than by covalent linkage, DNA

sequences encoding a DNA-binding domain, with any introduced sequence alterations, is joined to DNA encoding one or more suitably engineered ligand-binding domains, and if desired, to DNA encoding a transcriptional activation domain or other optional domain(s). These sequences are joined such that they constitute a single open reading frame that can be translated in cells into a single polypeptide harboring all component domains. The order and arrangement of the domains within the polypeptide can vary. At least two such chimeras are required for the optimal embodiment of this method. These constructions encode polypeptides containing distinct DNA-binding domains, ligand-binding domains with distinct specificity for multimerizing moieties, and in some embodiments, transcriptional activation domains with different properties. For example, this invention includes chimeras of the following structure:

(immunophilin) --- (txn activator) -- (DNA binding domain)

wherein "immunophilin" represents 1, 2 or 3 immunophilin domains, such as the FKBP12 domain of Spencer et al, "txn activator" represents a VP16 domain and "DNA binding domain" represents a DNA binding domain of Phox1 or SRE-ZBP.

As such, Gilman does not teach or suggest the claimed complexes in which the ligand mediates heterodimerization by binding to the DNA-binding polypeptide.

It is well-established that in order to be available as a reference under 35 U.S.C. § 102/103, the reference must contain an enabling disclosure. *See, e.g., Chester v. Miller*, 906 F.2d at 1576 n.2, 15 USPQ2d at 1336 n.2 (Fed. Cir. 1990); *Titanium Metals Corp. of America v. Banner*, 778 F.2d at 781, 227 USPQ at 778 (Fed. Cir. 1985); *Scripps Clinic & Research Found. v. Genentech, Inc.*, 927 F.2d 1565, 1578, 18 USPQ2d 1001, 1011 (Fed. Cir. 1991); *Helifix Ltd. v. Blok-Lok Ltd.*, 208 F.3d 1339, 54 USPQ2d 1299 (Fed. Cir. 2000). In other words, the reference must "sufficiently describe the claimed invention to have placed the public in possession of it." *See, Minnesota Mining & Mfg. Co. ("3M") v. Johnson & Johnson Orthopaedics, Inc.*, 976 F.2d 1559, 1572, 24 USPQ2d 1321, 1332 (Fed. Cir. 1992); *see also In re Donohue*, 766 F.2d 531, 533, 226 USPQ 619, 621 (Fed. Cir. 1985).

In the instant case, Gilman does not place the public in possession of ligand-mediated heterodimeric complexes as claimed. As noted throughout prosecution and above, Gilman discloses only complexes in which DNA-binding domains are covalently linked or in which

additional immunophilin ligand-binding domains are fused to the DNA-binding domain to mediate dimerization. This is in stark contrast to the claimed complexes in which the ligand mediates heterodimerization by binding to the DNA-binding domains. See, e.g., Example 1.3 on page 89 of the as-filed specification. The fact that the present applicants subsequently demonstrated complexes as claimed cannot be used to supplement the reference.


When taken as a whole, Gilman does not describe, demonstrate or in any way suggest complexes as claimed in claims 34 and 48. Since this reference does not place the public in possession of the complexes comprising a non-naturally occurring Cys2-His2 zinc finger protein bound via a ligand to a second DNA-binding domain, withdrawal of this rejection is in order.

CONCLUSION

For the reasons stated above, Appellants respectfully submit that the pending claims are novel and non-obvious. Accordingly, Appellants request that the rejections of the claims on appeal be reversed, and that the application be remanded to the Examiner so that the appealed claims can proceed to allowance.

Respectfully submitted,

Date: April 1, 2009

By: 
Dahna S. Pasternak
Registration No. 41,411
Attorney for Appellants

ROBINS & PASTERNAK LLP
1731 Embarcadero Road, Suite 230
Palo Alto, CA 94303
Telephone: (650) 493-3400
Facsimile: (650) 493-3440

CLAIMS APPENDIX

The claims on appeal are as follows:

34. A complex comprising:
- (a) a heterodimer comprising
 - (i) a first polypeptide, and
 - (ii) a second polypeptide; and
 - (b) a ligand that binds to the first and second polypeptides and mediates heterodimerization of the first and second polypeptides,
- wherein the first and second polypeptides bind to DNA, and further wherein the first or second polypeptide comprises an engineered, non-naturally occurring Cys2-His2 zinc finger binding domain.
48. A switching system comprising a protein switch comprising: (i) a first component comprising a first polypeptide and (ii) a second component comprising a second polypeptide, in which the first polypeptide binds to the second polypeptide, wherein binding of the first polypeptide to the second polypeptide forms a heterodimer and the binding of the first and second polypeptides is mediated by binding of a ligand to the first and second polypeptides, and (iii) a third component comprising the ligand, wherein the first and second polypeptides bind to DNA, and further wherein the first or second polypeptide comprises an engineered, non-naturally occurring Cys2-His2 zinc finger binding domain.

EVIDENCE APPENDIX

The following documents are attached to this Brief:

(1) a copy of *Ex parte Dewis* (2007) Appeal 2007-1610 (BPAI). This case was cited in the Response After Final mailed August 4, 2008. Expedited procedure was in order but an Advisory Action indicating consideration of this document was never received;

(2) WO 96/06166 by Medical Research Council, published February 29, 1996. This reference was cited as reference B6 in the IDS mailed on October 24, 2003 and was indicated considered by the Office by return of the initialed 1449s on May 24, 2004;

(3) WO 98/53057 by Medical Research Council, published November 26, 1998. This reference was cited as reference B6 in the IDS mailed on October 24, 2003 and was indicated considered by the Office by return of the initialed 1449s on May 24, 2004;

(4) WO 00/73434 by Gendaq Limited, published December 7, 2000. This reference was cited as reference B6 in the IDS mailed on October 24, 2003 and was indicated considered by the Office by return of the initialed 1449s on May 24, 2004.

RELATED PROCEEDINGS APPENDIX

As noted above on page 2 of this Appeal Brief, Applicants are not aware of any related, currently pending appeals or interferences. Accordingly, no documents are submitted with this Appendix.

The opinion in support of the decision being entered today
is *not* binding precedent of the Board.

UNITED STATES PATENT AND TRADEMARK OFFICE

**BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES**

Ex parte MARK LAWRENCE DEWIS,
DAVID JOHN EDWARDS, LESLEY KENDRICK,
MARIA WRIGHT, and AMIR YUSUF

Appeal 2007-1610
Application 10/955,833
Technology Center 1600

Decided: September 4, 2007

Before TONI R. SCHEINER, LORA M. GREEN, and RICHARD M.
LEBOVITZ, *Administrative Patent Judges*.

LEBOVITZ, *Administrative Patent Judge*.

DECISION ON APPEAL

This is a decision on appeal from the final rejection of claims 7-12.
We have jurisdiction under 35 U.S.C. § 6(b). We affirm.

STATEMENT OF CASE

A problem with developing flavoring agents for fruity
and herbaceous materials, such as mango flavor, is that natural
plant materials do not contain a single flavoring agent, but
rather contain a complex mixture of volatile components
making identification of characteristic flavors very difficult.

The volatiles of mango were analyzed by gas chromatography and a combined gas chromatograph-mass spectrometer. The volatiles were also analyzed by gas chromatography on a sulfur detector.

(Spec. 2: 21-27).

The Specification describes the discovery that ethyl 3-mercaptopbutyrate – identified from mango – can be used as a flavoring and perfuming agent because of its unique flavor and odorant properties (Spec. 1-2). The claims are drawn to an ingestible composition comprising an ingestible vehicle and ethyl 3-mercaptopbutyrate.

The following rejections are on appeal in this proceeding:

- 1) Claims 7-12 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement (Answer 13);
- 2) Claims 7-12 stand rejected (three separate rejections: of claims 7-12, 10-12, and 7; Answer 7, 9, and 13, respectively) under 35 U.S.C. § 112, second paragraph, as indefinite;
- 3) Claims 7-9 stand rejected under 35 U.S.C. § 102 as anticipated by Nielsen (“Stereoselective Reduction of Thiocarbonyl Compounds with Baker’s Yeast,” *Tetrahedron: Asymmetry*, 5: 403-410, 1994; referred to by the Examiner as “Nielsen and Madsen”) (Answer 11); and
- 4) Claim 7 stands rejected under 35 U.S.C. § 102(b) as anticipated by Lazier (US 2,402,639, issued Jun. 25, 1946; referred to by the Examiner as “Lazier and Signaigo”) (Answer 12).

The claims in each rejection stand or fall together because separate reasons for patentability were not provided for any individual claim. We select claims 7 and 10 as representative for deciding all rejections in this appeal. See 37 C.F.R. § 41.37(c)(1)(vii). Claims 7 and 10 read as follows:

7. An ingestible composition comprising:
(i) an ingestible vehicle; and
(ii) an organoleptically effective amount of ethyl 3-mercaptoputyrate represented by the formula,
 $\text{CH}_3(\text{SH})\text{CHCH}_2\text{COOCH}_2\text{CH}_3$ provided that the ethyl 3-mercaptoputyrate is not part of a naturally occurring mixture of compounds or part of a synthetic mixture of compounds which is the same as the naturally occurring mixture of compounds.

10. The ingestible composition according to claim 7, wherein the ingestible composition is a beverage product.

CLAIM INTERPRETATION

Claim 7 is drawn to an ingestible composition comprising (i) an ingestible vehicle and (ii) ethyl 3-mercaptoputyrate “provided that the ethyl 3-mercaptoputyrate is not part of a naturally occurring mixture of compounds or a part of a synthetic mixture of compounds which is the same as the naturally occurring mixture of compounds.”

At issue in this appeal is the proper interpretation of “provided that the ethyl 3-mercaptoputyrate is not part of a naturally occurring mixture of compounds.” We give the words in a claim their broadest reasonable interpretation as they would be understood by persons of skill in the art in the context of the Specification. *See In re Morris*, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997). In this case, the phrase “naturally occurring mixture of compounds” does not appear in the Specification as originally filed. However, “naturally occurring” would be understood by persons of skill in the art to mean that it exists or is found in nature – that is, it is “a product of nature” and not “a product of human ingenuity.” *Diamond v. Chakrabarty*, 447 US 303, 309, 313 (1980). Thus, we interpret a

“naturally occurring mixture of compounds” to mean a “mixture of compounds” that can be found in nature.

Ethyl 3-mercaptoputyrate was identified by the inventors as a flavorant present in the “complex mixture” of components that naturally occur in mango (Spec. 2: 21-27 and 5: 33 to 6:12). In this context, we interpret “provided that the ethyl 3-mercaptoputyrate is not part of a naturally occurring mixture of compounds” to mean that the mercaptoputyrate compound is not present in the claimed composition in the same complex form in which it would occur in nature.

We have considered, but reject, the Examiner’s alternative interpretation (Answer 6-7). As we understand it, the Examiner interprets “naturally occurring mixture of compounds” phrase to mean “a mixture of naturally occurring compounds.” In our opinion, the Examiner improperly interpreted “naturally occurring” to describe the compounds present in the mixture, rather than the entire mixture, itself.

The term “ingestible” as recited in claim 7 is also at issue in this proceeding. The Specification states the ethyl 3-mercaptoputyrate is useful for imparting a unique flavor to foodstuffs (Spec. 5: 33-35). It is described as useful “in a wide variety of ingestible vehicles” that include gum, confectionary products, and beverages (Spec. 8: 7-14). The term “ingestible” is also defined in the Specification to mean “all materials and compositions which are used by or which perform a function in the body” (Spec. 6: 17-21). Thus, we interpret the phrases “ingestible composition” and “ingestible vehicle” as recited in claim 7 to mean materials and compositions suitable as foods.

Written description rejection

Claims 7-12 stand rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. The Examiner contends that the phrase “provided that the ethyl 3-mercaptopbutyrate is not part of a naturally occurring mixture of compounds or part of a synthetic mixture of compounds which is the same as the naturally occurring mixture” of compounds is “new matter” to the application because it is not supported in the Specification as originally filed (Answer 13). “[N]owhere in the written description is language reflecting the present form of claim 7 found” (Final Office Action 9).

“The purpose of the written description requirement is to prevent an applicant from later asserting that he invented that which he did not; the applicant for a patent is therefore required ‘to recount his invention in such detail that his future claims can be determined to be encompassed within his original creation.’” *Amgen Inc. v. Hoechst Marion Roussel Inc.*, 314 F.3d 1313, 1330 [65 USPQ2d 1385] (Fed. Cir. 2003) (citing *Vas-Cath Inc. v. Mahurkar*, 935 F.2d 1555, 1561 [19 USPQ2d 1111] (Fed. Cir. 1991)).

While there is no requirement that the claimed invention be described in the identical wording that was used in the Specification, there must be sufficient disclosure to show one of skill in this art that the inventor “invented what is claimed.” See *Union Oil Co. of California v. Atlantic Richfield Co.*, 208 F.3d 989, 997, 54 USPQ2d 1227, 1235 (Fed. Cir. 2000).

According to the Specification, Appellants discovered that ethyl 3-mercaptopbutyrate “possesses unexpected flavor properties and imparts a unique note to flavors” especially in foodstuffs (Spec. 5: 33-37). It is present among “[a] relatively large number of components . . . identified in

an analysis of [a solvent extract of] mango” (Spec. 5: 37 to 38). Ethyl 3-mercaptopbutyrate is stated to be “present at such low concentrations in mango that it cannot be isolated from the fruit in a commercially viable way” (Spec. 6: 10-12). Instead, Appellants describe the chemical synthesis of ethyl 3-mercaptopbutyrate in a “purified form, unaccompanied by substances of natural origin present in mango” (Spec. 4: 35 to 5: 2) and shows that it acts as a beneficial flavorant (Spec. 38-39 (Example 2)). Thus, Appellants’ invention is the discovery that purified ethyl 3-mercaptopbutyrate acts as a flavoring when introduced into foodstuffs.

The written description must be of sufficient detail to show possession of the full scope of the invention. *Pandrol USA LP v. Airboss Railway Products Inc.*, 424 F.3d 1161, 1165, 76 USPQ2d 1524, 1527 (Fed. Cir. 2005). In this case, naturally occurring mixtures are excluded from the claims, but that leaves the claim open to everything else that contains ethyl 3-mercaptopbutyrate – including any composition, however modified that it is no longer naturally occurring.¹ In our opinion, such a claim scope is not justified nor drawn to what Appellants invented. The invention described in the Specification is “purified” ethyl 3-mercaptopbutyrate “unaccompanied by substances of natural origin present in mango” (Spec. 4: 35 to 5: 2) as a novel flavoring or perfuming agent. This is the only invention described in the Specification. There is no detail in the Specification that shows that Appellants possessed compositions of a different scope, let alone of an intermediate scope to cover mixtures of less complexity than the naturally-

¹ Such compositions would include, for example, less complex compositions derived from naturally-occurring mixtures by fractionation, extraction, and other processing steps.

occurring mixture from which ethyl 3-mercaptoputyrate was originally identified.

Granted, the purified ethyl 3-mercaptoputyrate described in the application is “not a part of a naturally occurring mixture of compounds.” However, what Appellants invented is a “purified” compound that, when introduced into a foodstuff, imparts a unique flavor to it. The only disclosure with respect to naturally occurring mixtures is that the concentration of ethyl 3-mercaptoputyrate is too low for it to be isolated from mango (Spec. 6: 10-12). As a consequence, ethyl 3-mercaptoputyrate was chemically synthesized – the form which is characterized in the Specification as “purified.” In sum, we agree with the Examiner that claim 7 lacks a written description in the application.

Our decision is consistent with *In re Johnson and Farnham*, 558 F.2d 1008, 194 USPQ 187 (CCPA 1977), a CCPA case which dealt with exclusionary language in a claim that was not present in the application upon which priority was based. In *Johnson*, the applicant was attempting to narrow the scope of a claimed genus of compounds by excluding two species which had been lost in an interference. The Examiner, in a rejection affirmed by the Board of Appeals, asserted that the claims were not entitled to the 1963 filing date of the application because the claimed subject matter was not described in it as required by 35 U.S.C. § 112, first paragraph. The CCPA reversed. “The only inquiry is whether, after exclusion from the original claims of two species specifically disclosed in the 1963 application, the 1963 disclosure satisfies § 112, first paragraph, for the ‘limited’ genus now claimed.” *Johnson*, 558 F.2d at 1017-1018, 194 USPQ at 195.

The CCPA found that it did because its priority application contained “a broad and complete generic disclosure, coupled with extensive examples fully supportive of the limited genus now claimed.” *Johnson*, 558 F.2d at 1018, 194 USPQ at 196.

The CCPA distinguished an earlier case, *Welstead*, in which an applicant sought to exclude subject matter from an originally claimed genus, because in that case the new subgenus was not described in the application nor was there a description of “[its] species thereof amounting in the aggregate to the same thing.” *Johnson*, 558 F.2d at 1018, 194 USPQ at 196.

The CCPA concluded:

The notion that one who fully discloses, and teaches those skilled in the art how to make and use, a genus and numerous species therewithin, has somehow failed to disclose, and teach those skilled in the art how to make and use, that genus minus two of those species, and has thus failed to satisfy the requirements of § 112, first paragraph, appears to result from a hypertechnical application of legalistic prose relating to that provision of the statute.

Johnson, 558 F.2d at 1019, 194 USPQ at 196.

In this case, there is no description in the Specification – as there was in *Johnson* – of a genus minus what has been excluded from the claim. The Specification describes only one species – purified ethyl 3-mercaptoputyrate – and no other. There is no detailed description to show that Appellants possessed the invention which is now claimed.

Appellants argue that “[i]t has always been clear that appellant merely wishes to claim ethyl 3-mercaptoputyrate in purified form as an organoleptic agent and not ethyl 3-mercaptoputyrate in a naturally occurring mixture of compounds or part of a synthetic mixture of compounds which is the same

Appeal 2007-1610
Application 10/955,833

as the naturally occurring mixture of compounds” (Br. 11). However, purified claim ethyl 3-mercaptoputyrate is not what is presently claimed.

Thus, we conclude that the phrase “provided that the ethyl 3-mercaptoputyrate is not part of a naturally occurring mixture of compounds or a part of a synthetic mixture of compounds which is the same as the naturally occurring mixture of compounds” is new matter to the Specification in violation of the written description requirement of 35 U.S.C. § 112, first paragraph. The rejection of claims 7-12 is affirmed.

Indefiniteness rejection under § 112, second paragraph

There are three rejections at issue in this appeal for lack of definiteness under 35 U.S.C. § 112, second paragraph. First, claims 7-12 stand rejected as indefinite because “it is unclear exactly what constitutes, in the context of the invention, ‘a naturally occurring mixture of compounds.’” (Answer 7.) Related to this issue, the Examiner states that if the claims are interpreted to exclude any mixture of naturally occurring compounds, “the compositions specified in claims 10-12 lack antecedent basis” because they would exclude Appellants’ “most preferred embodiments: the beverage, confection and chewing gum” (Answer 9-10). Third, the Examiner states that claim 7 is indefinite “[b]ecause a naturally occurring mixture and a synthetic mixture are *not* the same, they cannot as a matter of fact properly be characterized as such” (Answer 13).

We reverse the rejections. The phrase “naturally occurring mixture of compounds,” when properly interpreted, means a “mixture of compounds” that can be found in nature (see *supra* at p. 3-4). This is not indefinite nor does it lead claims 10-12 to lack antecedent basis.

The characterization of the synthetic mixture as being the “same” as the naturally occurring mixture would be understood by persons of skill in the art to mean that the profile of compounds in the mixtures are the same. Thus, we do not find that this term introduced ambiguity into the claim.

Anticipation by Nielsen

Claims 7-9 stand rejected under 35 U.S.C. § 102 as anticipated by Nielsen.

Nielsen describes the synthesis of ethyl 3-mercaptoputyrate (Nielsen, at 408; Answer 11). The ethyl 3-mercaptoputyrate accumulates in a hexane phase in the reaction vessel (Nielsen, at 408; Answer 11). The Examiner contends that “[s]ince hexane is an ingestible vehicle, in the broadest reasonable interpretation of the term, when considered in light of the instant specification, the Nielsen . . . reference is anticipatory. Hexane is capable of being ingested, thus it is an ingestible material” (Answer 11).

Appellants contend that hexane is not an “ingestible vehicle” as would be understood in the light of the Specification (Br. 7-8). “As set out in appellant’s specification, ‘ingestible’ means to take in as food. Appellant’s specification states that ‘[a]pplicant has discovered that ethyl 3-mercaptoputyrate . . . possesses unexpected flavor properties and imparts a unique note to flavors, *especially for conferring in foodstuffs* . . .’ Appellant’s specification at page 5, lines 27-31. (emphasis added)” (Br. 8). Appellants provide evidence that hexane is “a toxic substance causing central nervous system effects including dizziness, giddiness, nausea, and headache” and therefore not ingestible as a food (Br. 7-8).

In our opinion, Appellants have the better argument. Claim terms are given their broadest reasonable interpretation as they would be understood by persons of ordinary skill in the art when read in the context of the Specification. We have interpreted “ingestible” to mean a material that can be present in a food (see *supra* at p. 4) because the Specification describes the invention as purified ethyl 3-mercaptoputyrate as a flavoring to be used in foodstuffs (Spec. 5: 33-38). The Examiner’s interpretation of “ingestible vehicle” is broad, but not *reasonable* in light of the Specification’s teaching about the use of ethyl 3-mercaptoputyrate in food.

Appellants have introduced evidence, un rebutted by the Examiner, that hexane is a toxic substance and therefore would not be considered an “ingestible vehicle” as required by claim 7. We find this evidence persuasive, and thus concur with Appellants that the Examiner erred in rejecting claims 7-9 as anticipated by Nielsen. We reverse this rejection.

Anticipation by Lazier

Claim 7 stands rejected under 35 U.S.C. § 102(b) as anticipated by Lazier.

Lazier teaches the synthesis of ethyl 3-mercaptoputyrate having 87% purity (Lazier, at col. 3, ll. 35-37; Answer 12). The Examiner contends that this composition meets the limitation of claim 7 requiring the presence of an ingestible vehicle “because there is some additional material contained besides the mercapto-ester compound (the ‘ingestible vehicle’)” (Answer 12).

Appellants contend that “[t]he Examiner may NOT assume that this additional material (13%) is an ingestible material. Lazier et al. does not

identify this additional material. This additional material could just as readily be one or more toxic (non-food) substances. Lazier et al. was not seeking to make flavoring agents for use in ingestible vehicles but rather was seeking to make starting materials for use in polymers (Lazier et al. at col. 1, lines 4-9). Hence, Lazier et al. was not concerned whether this additional material (13%) was an ingestible material" (Br. 10).

"A patent is invalid for anticipation if a single prior art reference discloses each and every limitation of the claimed invention. Moreover, a prior art reference may anticipate without disclosing a feature of the claimed invention if that missing characteristic is necessarily present, or inherent, in the single anticipating reference." *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1377, 67 USPQ2d 1664, 1667 (Fed. Cir. 2003) (internal citations omitted). See also *SmithKline Beecham Corp. v. Apotex Corp.*, 403 F.3d 1331, 1343 74 USPQ2d 1398, 1406 (Fed. Cir. 2005). "[W]hen the PTO shows sound basis for believing that the products of the applicant and the prior art are the same, the applicant has the burden of showing that they are not." *In re Spada*, 911 F.2d 705, 708, 15 USPQ2d 1655, 1658 (Fed. Cir. 1990).

The issue raised by this rejection is whether the Examiner has provided a reasonable basis for shifting the burden to Appellants to establish that the claimed composition is distinguishable from Lazier's composition; and if so, whether Appellants' burden has been met. In our opinion, the Examiner met his burden, but Appellants did not.

Lazier's Example II, relied upon by the Examiner for its disclosure of a fraction that "analyzes for 87% purity as ethyl 3-mercaptoputyrate" (Lazier, at col. 3, ll. 36-38), also comprises "[w]ater . . . formed in the course

Appeal 2007-1610
Application 10/955,833

of the reaction” (Lazier, at col. 3, ll. 38-39). Since water is an ingestible vehicle, we conclude that its presence is enough to provide reasonable basis for considering Lazier’s composition to be the same as the composition of claim 7. Appellants had the opportunity to provide evidence that Lazier’s synthetic method would not result in an ingestible composition as required by claim 7, but no evidence was offered in rebuttal. Accordingly, we affirm the rejection.

TIME PERIOD

No time period for taking any subsequent action in connection with this appeal may be extended under 37 CFR § 1.136(a).

AFFIRMED

Ssc

RICHARD R. MUCCINO
758 SPRINGFIELD AVENUE
SUMMIT, NJ 07901



INTERNATIONAL APPLICATION PUBLISHED UNDER

WO 9606166A1

(51) International Patent Classification⁶:C12N 15/10, 15/12, 15/62, C12Q 1/68,
C07K 14/47, A61K 48/00

A1

(11) International Publication Number:

WO 96/06166

(43) International Publication Date:

29 February 1996 (29.02.96)

(21) International Application Number: PCT/GB95/01949

(22) International Filing Date: 17 August 1995 (17.08.95)

(30) Priority Data:

9416880.4	20 August 1994 (20.08.94)	GB
9422534.9	8 November 1994 (08.11.94)	GB
9514698.1	18 July 1995 (18.07.95)	GB

(71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): CHOO, Yen [SG/SG]; Alexandra Park, 5 Hyderabad Road, Singapore 0511 (SG). KLUG, Aaron [GB/GB]; 70 Cavendish Avenue, Cambridge CB1 4UT (GB). GARCIA, Isidro-Sanchez [ES/ES]; Cuesta del Sancti-Spiritus, 6-8, 5^oD, E-37001 Salamanca (ES).

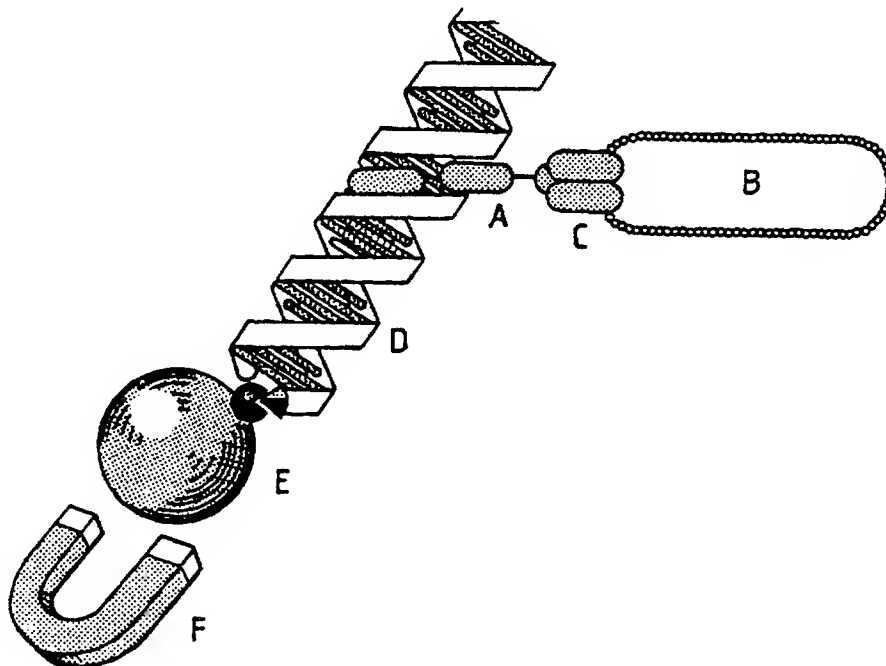
(74) Agent: KEITH W. NASH & CO.; Pearl Assurance House, 90-92 Regent Street, Cambridge CB2 1DP (GB).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

(54) Title: IMPROVEMENTS IN OR RELATING TO BINDING PROTEINS FOR RECOGNITION OF DNA



(57) Abstract

Disclosed are libraries of DNA sequences encoding zinc finger binding motifs for display on a particle, together with methods of designing zinc finger binding polypeptides for binding to a particular target sequence and, inter alia, use of designed zinc finger polypeptides for various *in vitro* or *in vivo* applications.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

Title: Improvements in or Relating to Binding Proteins for Recognition of DNA

Field of the Invention

This invention relates *inter alia* to methods of selecting and designing polypeptides comprising zinc finger binding motifs, polypeptides made by the method(s) of the invention and to various applications thereof.

Background of the Invention

Selective gene expression is mediated via the interaction of protein transcription factors with specific nucleotide sequences within the regulatory region of the gene. The most widely used domain within protein transcription factors appears to be the zinc finger (Zf) motif. This is an independently folded zinc-containing mini-domain which is used in a modular repeating fashion to achieve sequence-specific recognition of DNA (Klug 1993 Gene 135, 83-92). The first zinc finger motif was identified in the *Xenopus* transcription factor TFIIA (Miller *et al.*, 1985 EMBO J. 4, 1609-1614). The structure of Zf proteins has been determined by NMR studies (Lee *et al.*, 1989 Science 245, 635-637) and crystallography (Pavletich & Pabo, 1991 Science 252, 809-812).

The manner in which DNA-binding protein domains are able to discriminate between different DNA sequences is an important question in understanding crucial processes such as the control of gene expression in differentiation and development. The zinc finger motif has been studied extensively, with a view to providing some insight into this problem, owing to its remarkable prevalence in the eukaryotic genome, and its important role in proteins which control gene expression in *Drosophila* (e.g. Harrison & Travers 1990 EMBO J. 9, 207-216), the mouse (Christy *et al.*, 1988 Proc. Natl. Acad. Sci. USA 85, 7857-7861) and humans (Kinzler *et al.*, 1988 Nature (London) 332, 371).

Most sequence-specific DNA-binding proteins bind to the DNA double helix by inserting an α -helix into the major groove (Pabo & Sauer 1992 Annu. Rev. Biochem. 61,

1053-1095; Harrison 1991 Nature (London) 353, 715-719; and Klug 1993 Gene 135, 83-92). Sequence specificity results from the geometrical and chemical complementarity between the amino acid side chains of the α -helix and the accessible groups exposed on the edges of base-pairs. In addition to this direct reading of the DNA sequence, interactions with the DNA backbone stabilise the complex and are sensitive to the conformation of the nucleic acid, which in turn depends on the base sequence (Dickerson & Drew 1981 J. Mol. Biol. 149, 761-786). *A priori*, a simple set of rules might suffice to explain the specific association of protein and DNA in all complexes, based on the possibility that certain amino acid side chains have preferences for particular base-pairs. However, crystal structures of protein-DNA complexes have shown that proteins can be idiosyncratic in their mode of DNA recognition, at least partly because they may use alternative geometries to present their sensory α -helices to DNA, allowing a variety of different base contacts to be made by a single amino acid and vice versa (Matthews 1988 Nature (London) 335, 294-295).

Mutagenesis of Zf proteins has confirmed modularity of the domains. Site directed mutagenesis has been used to change key Zf residues, identified through sequence homology alignment, and from the structural data, resulting in altered specificity of Zf domain (Nardelli *et al.*, 1992 NAR 20, 4137-4144). The authors suggested that although design of novel binding specificities would be desirable, design would need to take into account sequence and structural data. They state "there is no prospect of achieving a zinc finger recognition code".

Despite this, many groups have been trying to work towards such a code, although only limited rules have so far been proposed. For example, Desjarlais *et al.*, (1992b PNAS 89, 7345-7349) used systematic mutation of two of the three contact residues (based on consensus sequences) in finger two of the polypeptide Sp1 to suggest that a limited degenerate code might exist. Subsequently the authors used this to design three Zf proteins with different binding specificities and affinities (Desjarlais & Berg, 1993 PNAS 90, 2250-2260). They state that the design of Zf proteins with predictable specificities and affinities "may not always be straightforward".

We believe the zinc finger of the TFIIIA class to be a good candidate for deriving a set of more generally applicable specificity rules owing to its great simplicity of structure and interaction with DNA. The zinc finger is an independently folding domain which uses a zinc ion to stabilise the packing of an antiparallel β -sheet against an α -helix (Miller *et al.*, 1985 EMBO J. 4, 1609-1614; Berg 1988 Proc. Natl. Acad. Sci. USA 85, 99-102; and Lee *et al.*, 1989 Science 245, 635-637). The crystal structures of zinc finger-DNA complexes show a semiconserved pattern of interactions in which 3 amino acids from the α -helix contact 3 adjacent bases (a triplet) in DNA (Pavletich & Pabo 1991 Science 252, 809-817; Fairall *et al.*, 1993 Nature (London) 366, 483-487; and Pavletich & Pabo 1993 Science 261, 1701-1707). Thus the mode of DNA recognition is principally a one-to-one interaction between amino acids and bases. Because zinc fingers function as independent modules (Miller *et al.*, 1985 EMBO J. 4, 1609-1614; Klug & Rhodes 1987 Trends Biochem. Sci. 12, 464-469), it should be possible for fingers with different triplet specificities to be combined to give specific recognition of longer DNA sequences. Each finger is folded so that three amino acids are presented for binding to the DNA target sequence, although binding may be directly through only two of these positions. In the case of Zif268 for example, the protein is made up of three fingers which contact a 9 base pair contiguous sequence of target DNA. A linker sequence is found between fingers which appears to make no direct contact with the nucleic acid.

Protein engineering experiments have shown that it is possible to alter rationally the DNA-binding characteristics of individual zinc fingers when one or more of the α -helical positions is varied in a number of proteins (Nardelli *et al.*, 1991 Nature (London) 349, 175-178; Nardelli *et al.*, 1992 Nucleic Acids Res. 20, 4137-4144; and Desjarlais & Berg 1992a Proteins 13, 272). It has already been possible to propose some principles relating amino acids on the α -helix to corresponding bases in the bound DNA sequence (Desjarlais & Berg 1992b Proc. Natl. Acad. Sci. USA 89, 7345-7349). However in this approach the altered positions on the α -helix are prejudged, making it possible to overlook the role of positions which are not currently considered important; and secondly, owing to the importance of context, concomitant alterations are sometimes required to affect specificity (Desjarlais & Berg 1992b), so that a significant correlation between an amino acid and base may be misconstrued.

To investigate binding of mutant Zf proteins, Thiesen and Bach (1991 FEBS 283, 23-26) mutated Zf fingers and studied their binding to randomised oligonucleotides, using electrophoretic mobility shift assays. Subsequent use of phage display technology has permitted the expression of random libraries of Zf mutant proteins on the surface of bacteriophage. The three Zf domains of Zif268, with 4 positions within finger one randomised, have been displayed on the surface of filamentous phage by Rebar and Pabo (1994 Science 263, 671-673). The library was then subjected to rounds of affinity selection by binding to target DNA oligonucleotide sequences in order to obtain Zf proteins with new binding specificities. Randomised mutagenesis (at the same positions as those selected by Rebar & Pabo) of finger 1 of Zif 268 with phage display has also been used by Jamieson *et al.*, (1994 Biochemistry 33, 5689-5695) to create novel binding specificity and affinity.

More recently Wu *et al.* (1995 Proc. Natl. Acad. Sci. USA 92, 344-348) have made three libraries, each of a different finger from Zif268, and each having six or seven α -helical positions randomised. Six triplets were used in selections but did not return fingers with any sequence biases; and when the three triplets of the Zif268 binding site were individually used as controls, the vast majority of selected fingers did not resemble the sequences of the wild-type Zif268 fingers and, though capable of tight binding to their target sites *in vitro*, were usually not able to discriminate strongly against different triplets. The authors interpret the results as evidence against the existence of a code.

In summary, it is known that Zf protein motifs are widespread in DNA binding proteins and that binding is via three key amino acids, each one contacting a single base pair in the target DNA sequence. Motifs are modular and may be linked together to form a set of fingers which recognise a contiguous DNA sequence (e.g. a three fingered protein will recognise a 9mer etc). The key residues involved in DNA binding have been identified through sequence data and from structural information. Directed and random mutagenesis has confirmed the role of these amino acids in determining specificity and affinity. Phage display has been used to screen for new binding specificities of random mutants of fingers. A recognition code, to aid design of new finger specificities, has been worked towards although it has been suggested that specificity may be difficult to predict.

Summary of the Invention

In a first aspect the invention provides a library of DNA sequences, each sequence encoding at least one zinc finger binding motif for display on a viral particle, the sequences coding for zinc finger binding motifs having random allocation of amino acids at positions -1, +2, +3, +6 and at least at one of positions +1, +5 and +8.

A zinc finger binding motif is the α -helical structural motif found in zinc finger binding proteins, well known to those skilled in the art. The above numbering is based on the first amino acid in the α -helix of the zinc finger binding motif being position +1. It will be apparent to those skilled in the art that the amino acid residue at position -1 does not, strictly speaking, form part of the α -helix of the zinc binding finger motif. Nevertheless, the residue at -1 is shown to be very important functionally and is therefore considered as part of the binding motif α -helix for the purposes of the present invention.

The sequences may code for zinc finger binding motifs having random allocation at all of positions +1, +5 and +8. The sequences may also be randomised at other positions (e.g. at position +9, although it is generally preferred to retain an arginine or a lysine residue at this position). Further, whilst allocation of amino acids at the designated "random" positions may be genuinely random, it is preferred to avoid a hydrophobic residue (Phe, Trp or Tyr) or a cysteine residue at such positions.

Preferably the zinc finger binding motif is present within the context of other amino acids (which may be present in zinc finger proteins), so as to form a zinc finger (which includes an antiparallel β -sheet). Further, the zinc finger is preferably displayed as part of a zinc finger polypeptide, which polypeptide comprises a plurality of zinc fingers joined by an intervening linker peptide. Typically the library of sequences is such that the zinc finger polypeptide will comprise two or more zinc fingers of defined amino acid sequence (generally the wild type sequence) and one zinc finger having a zinc finger binding motif randomised in the manner defined above. It is preferred that the randomised finger of the polypeptide is positioned between the two or more fingers having defined sequence. The

defined fingers will establish the "phase" of binding of the polypeptide to DNA, which helps to increase the binding specificity of the randomised finger.

Preferably the sequences encode the randomised binding motif of the middle finger of the Zif268 polypeptide. Conveniently, the sequences also encode those amino acids N-terminal and C-terminal of the middle finger in wild type Zif268, which encode the first and third zinc fingers respectively. In a particular embodiment, the sequence encodes the whole of the Zif268 polypeptide. Those skilled in the art will appreciate that alterations may also be made to the sequence of the linker peptide and/or the β -sheet of the zinc finger polypeptide.

In a further aspect, the invention provides a library of DNA sequences, each sequence encoding the zinc finger binding motif of at least a middle finger of a zinc finger binding polypeptide for display on a viral particle, the sequences coding for the binding motif having random allocation of amino acids at positions -1, +2, +3 and +6. Conveniently, the zinc finger polypeptide will be Zif268.

Typically, the sequences of either library are such that the zinc finger binding domain can be cloned as a fusion with the minor coat protein (pIII) of bacteriophage fd. Conveniently, the encoded polypeptide includes the tripeptide sequence Met-Ala-Glu as the N terminal of the zinc finger domain, which is known to allow expression and display using the bacteriophage fd system. Desirably the library comprises 10^6 or more different sequences (ideally, as many as is practicable).

In another aspect the invention provides a method of designing a zinc finger polypeptide for binding to a particular target DNA sequence, comprising screening each of a plurality of zinc finger binding motifs against at least an effective portion of the target DNA sequence, and selecting those motifs which bind to the target DNA sequence. An effective portion of the target DNA sequence is a sufficient length of DNA to allow binding of the zinc binding motif to the DNA. This is the minimum sequence information (concerning the target DNA sequence) that is required. Desirably at least two, preferably three or more, rounds of screening are performed.

The invention also provides a method of designing a zinc finger polypeptide for binding to a particular target DNA sequence, comprising comparing the binding of each of a plurality of zinc finger binding motifs to one or more DNA triplets, and selecting those motifs exhibiting preferable binding characteristics. Preferably the method defined immediately above is preceded by a screening step according to the method defined in the previous paragraph.

It is thus preferred that there is a two-step selection procedure: the first step comprising screening each of a plurality of zinc finger binding motifs (typically in the form of a display library), mainly or wholly on the basis of affinity for the target sequence; the second step comprising comparing binding characteristics of those motifs selected by the initial screening step, and selecting those having preferable binding characteristics for a particular DNA triplet.

Where the plurality of zinc finger binding motifs is screened against a single DNA triplet, it is preferred that the triplet is represented in the target DNA sequence at the appropriate position. However, it is also desirable to compare the binding of the plurality of zinc binding motifs to one or more DNA triplets not represented in the target DNA sequence (e.g. differing by just one of the three base pairs) in order to compare the specificity of binding of the various binding motifs. The plurality of zinc finger binding motifs may be screened against all 64 possible permutations of 3 DNA bases.

Once suitable zinc finger binding motifs have been identified and obtained, they will advantageously be combined in a single zinc finger polypeptide. Typically this will be accomplished by use of recombinant DNA technology; conveniently a phage display system may be used.

In another aspect, the invention provides a DNA library consisting of 64 sequences, each sequence comprising a different one of the 64 possible permutations of three DNA bases in a form suitable for use in the selection method defined above. Desirably the sequences are associated, or capable of being associated, with separation means. Advantageously, the separation means is selected from one of the following: microtitre plate; magnetic

beads; or affinity chromatography column. Conveniently the sequences are biotinylated. Preferably the sequences are contained within 12 mini-libraries, as explained elsewhere.

In a further aspect the invention provides a zinc finger polypeptide designed by one or both of the methods defined above. Preferably the zinc finger polypeptide designed by the method comprises a combination of a plurality of zinc fingers (adjacent zinc fingers being joined by an intervening linker peptide), each finger comprising a zinc finger binding motif. Desirably, each zinc finger binding motif in the zinc finger polypeptide has been selected for preferable binding characteristics by the method defined above. The intervening linker peptide may be the same between each adjacent zinc finger or, alternatively, the same zinc finger polypeptide may contain a number of different linker peptides. The intervening linker peptide may be one that is present in naturally-occurring zinc finger polypeptides or may be an artificial sequence. In particular, the sequence of the intervening linker peptide may be varied, for example, to optimise binding of the zinc finger polypeptide to the target sequence.

Where the zinc finger polypeptide comprises a plurality of zinc binding motifs, it is preferred that each motif binds to those DNA triplets which represent contiguous or substantially contiguous DNA in the sequence of interest. Where several candidate binding motifs or candidate combinations of motifs exist, these may be screened against the actual target sequence to determine the optimum composition of the polypeptide. Competitor DNA may be included in the screening assay for comparison, as described below.

The non-specific component of all protein-DNA interactions, which includes contacts to the sugar-phosphate backbone as well as ambiguous contacts to base-pairs, is a considerable driving force towards complex formation and can result in the selection of DNA-binding proteins with reasonable affinity but without specificity for a given DNA sequence. Therefore, in order to minimise these non-specific interactions when designing a polypeptide, selections should preferably be performed with low concentrations of specific binding site in a background of competitor DNA, and binding should desirably take place in solution to avoid local concentration effects and the avidity of multivalent

phage for ligands immobilised on solid surfaces.

As a safeguard against spurious selections, the specificity of individual phage should be determined following the final round of selection. Instead of testing for binding to a small number of binding sites, it would be desirable to screen all possible DNA sequences.

It has now been shown possible by the present inventors (below) to design a truly modular zinc binding polypeptide, wherein the zinc binding motif of each zinc binding finger is selected on the basis of its affinity for a particular triplet. Accordingly, it should be well within the capability of one of normal skill in the art to design a zinc finger polypeptide capable of binding to any desired target DNA sequence simply by considering the sequence of triplets present in the target DNA and combining in the appropriate order zinc fingers comprising zinc finger binding motifs having the necessary binding characteristics to bind thereto. The greater the length of known sequence of the target DNA, the greater the number of zinc finger binding motifs that can be included in the zinc finger polypeptide. For example, if the known sequence is only 9 bases long then three zinc finger binding motifs can be included in the polypeptide. If the known sequence is 27 bases long then, in theory, up to nine binding motifs could be included in the polypeptide. The longer the target DNA sequence, the lower the probability of its occurrence in any given portion of DNA.

Moreover, those motifs selected for inclusion in the polypeptide could be artificially modified (e.g. by directed mutagenesis) in order to optimise further their binding characteristics. Alternatively (or additionally) the length and amino acid sequence of the linker peptide joining adjacent zinc binding fingers could be varied, as outlined above. This may have the effect of altering the position of the zinc finger binding motif relative to the DNA sequence of interest, and thereby exert a further influence on binding characteristics.

Generally, it will be preferred to select those motifs having high affinity and high specificity for the target triplet.

In a further aspect, the invention provides a kit for making a zinc finger polypeptide for binding to a nucleic acid sequence of interest, comprising: a library of DNA sequences encoding zinc finger binding motifs of known binding characteristics in a form suitable for cloning into a vector; a vector molecule suitable for accepting one or more sequences from the library; and instructions for use.

Preferably the vector is capable of directing the expression of the cloned sequences as a single zinc finger polypeptide. In particular it is preferred that the vector is capable of directing the expression of the cloned sequences as a single zinc finger polypeptide displayed on the surface of a viral particle, typically of the sort of viral display particle which are known to those skilled in the art. The DNA sequences are preferably in such a form that the expressed polypeptides are capable of self-assembling into a number of zinc finger polypeptides.

It will be apparent that the kit defined above will be of particular use in designing a zinc finger polypeptide comprising a plurality of zinc finger binding motifs, the binding characteristics of which are already known. In another aspect the invention provides a kit for use when zinc finger binding motifs with suitable binding characteristics have not yet been identified, such that the invention provides a kit for making a zinc finger polypeptide for binding to a nucleic acid sequence of interest, comprising: a library of DNA sequences, each encoding a zinc finger binding motif in a form suitable for screening and/or selecting according to the methods defined above; and instructions for use.

Advantageously, the library of DNA sequences in the kit will be a library in accordance with the first aspect of the invention. Conveniently, the kit may also comprise a library of 64 DNA sequences, each sequence comprising a different one of the 64 possible permutations of three DNA bases, in a form suitable for use in the selection method defined previously. Typically, the 64 sequences are present in 12 separate mini-libraries, each mini-library having one position in the relevant triplet fixed and two positions randomised. Preferably, the kit will also comprise appropriate buffer solutions, and/or reagents for use in the detection of bound zinc fingers. The kit may also usefully include a vector suitable for accepting one or more sequences selected from the library of DNA

sequences encoding zinc finger binding motifs.

In a preferred embodiment, the present teaching will be used for isolating the genes for the middle zinc fingers which, having been previously selected by one of the 64 triplets, are thought to have specific DNA binding activity. The mixture of genes specifying fingers which bind to a given triplet will be amplified by PCR using three sets of primers. The sets will have unique restriction sites, which will define the assembly of zinc fingers into three finger polypeptides. The appropriate reagents are preferably provided in kit form.

For instance, the first set of primers might have *Sfi*I and *Age*I sites, the second set *Age*I and *Eag*I sites and third set *Eag*I and *Nor*I sites. It will be noted that the "first" site will preferably be *Sfi*I, and the "last" site *Nor*I, so as to facilitate cloning into the *Sfi*I and *Nor*I sites of the phage vector. To assemble a library of three finger proteins which recognise the sequence AAAGGGGGG, the fingers selected by the triplet GGG are amplified using the first two sets of primers and ligated to the fingers selected by the triplet AAA amplified using the third set of primers. The combinatorial library is cloned on the surface of phage and a nine base-pair site can be used to select the best combination of fingers en bloc.

The genes for fingers which bind to each of the 64 triplets can be amplified by each set of primers and cut using the appropriate restriction enzymes. These building blocks for three-finger proteins can be sold as components of a kit for use as described above. The same could be done for the library amplified with different primers so that 4- or 5- finger proteins could be built.

Additionally a large (pre-assembled) library of all combinations of the fingers selected by all triplets can also be developed for single-step selection of DNA-binding proteins using 9bp, or much longer, DNA fragments. For this particular application, which will require very large libraries of novel 3-finger proteins, it may be preferable to use methods of selection other than phage display; for example stalled polysomes (developed by Affimax) where protein and mRNA become linked.

In a further aspect the invention provides a method of altering the expression of a gene of interest in a target cell, comprising : determining (if necessary) at least part of the DNA sequence of the structural region and/or a regulatory region of the gene of interest; designing a zinc finger polypeptide to bind to the DNA of known sequence, and causing said zinc finger polypeptide to be present in the target cell, (preferably in the nucleus thereof). (It will be apparent that the DNA sequence need not be determined if it is already known.)

The regulatory region could be quite remote from the structural region of the gene of interest (e.g. a distant enhancer sequence or similar). Preferably the zinc finger polypeptide is designed by one or both of the methods of the invention defined above.

Binding of the zinc finger polypeptide to the target sequence may result in increased or reduced expression of the gene of interest depending, for example, on the nature of the target sequence (e.g. structural or regulatory) to which the polypeptide binds.

In addition, the zinc finger polypeptide may advantageously comprise functional domains from other proteins (e.g. catalytic domains from restriction enzymes, recombinases, replicases, integrases and the like) or even "synthetic" effector domains. The polypeptide may also comprise activation or processing signals, such as nuclear localisation signals. These are of particular usefulness in targeting the polypeptide to the nucleus of the cell in order to enhance the binding of the polypeptide to an intranuclear target (such as genomic DNA). A particular example of such a localisation signal is that from the large T antigen of SV40. Such other functional domains/signals and the like are conveniently present as a fusion with the zinc finger polypeptide. Other desirable fusion partners comprise immunoglobulins or fragments thereof (eg. Fab, scFv) having binding activity.

The zinc finger polypeptide may be synthesised *in situ* in the cell as a result of delivery to the cell of DNA directing expression of the polypeptide. Methods of facilitating delivery of DNA are well-known to those skilled in the art and include, for example, recombinant viral vectors (e.g. retroviruses, adenoviruses), liposomes and the like. Alternatively, the zinc finger polypeptide could be made outside the cell and then delivered

thereto. Delivery could be facilitated by incorporating the polypeptide into liposomes etc. or by attaching the polypeptide to a targeting moiety (such as the binding portion of an antibody or hormone molecule). Indeed, one significant advantage of zinc finger proteins over oligonucleotides or protein-nucleic acids (PNAs) in controlling gene expression, would be the vector-free delivery of protein to target cells. Unlike the above, many examples of soluble proteins entering cells are known, including antibodies to cell surface receptors. The present inventors are currently carrying out fusions of anti-bcr-abl fingers (see example 3 below) to a single-chain (sc) Fv fragment capable of recognising NIP (4-hydroxy-5-iodo-3-nitrophenyl acetyl). Mouse transferrin conjugated with NIP will be used to deliver the fingers to mouse cells via the mouse transferrin receptor.

Media (e.g. microtitre wells, resins etc.) coated with NIP can also be used as solid supports for zinc fingers fused to anti-NIP scFvs, for applications requiring immobilised zinc fingers (e.g. the purification of specific nucleic acids).

In a particular embodiment, the invention provides a method of inhibiting cell division by causing the presence in a cell of a zinc finger polypeptide which inhibits the expression of a gene enabling the cell to divide.

In a specific embodiment, the invention provides a method of treating a cancer, comprising delivering to a patient, or causing to be present therein, a zinc finger polypeptide which inhibits the expression of a gene enabling the cancer cells to divide. The target could be, for example, an oncogene or a normal gene which is overexpressed in the cancer cells.

To the best knowledge of the inventors, design of a zinc finger polypeptide and its successful use in modulation of gene expression (as described below) has never previously been demonstrated. This breakthrough presents numerous possibilities. In particular, zinc finger polypeptides could be designed for therapeutic and/or prophylactic use in regulating the expression of disease-associated genes. For example, zinc finger polypeptides could be used to inhibit the expression of foreign genes (e.g. the genes of bacterial or viral pathogens) in man or animals, or to modify the expression of mutated host genes (such

; oncogenes).

The invention therefore provides a zinc finger polypeptide capable of inhibiting the expression of a disease-associated gene. Typically the zinc finger polypeptide will not be a naturally-occurring polypeptide but will be specifically designed to inhibit the expression of the disease-associated gene. Conveniently the polypeptide will be designed by one or both of the methods of the invention defined above. Advantageously the disease-associated gene will be an oncogene, typically the *BCR-ABL* fusion oncogene or a ras oncogene. In a particular embodiment the invention provides a zinc finger polypeptide designed to bind to the DNA sequence GCAGAAGCC and capable of inhibiting the expression of the *BCR-ABL* fusion oncogene.

In yet another aspect the invention provides a method of modifying a nucleic acid sequence of interest present in a sample mixture by binding thereto a zinc finger polypeptide, comprising contacting the sample mixture with a zinc finger polypeptide having affinity for at least a portion of the sequence of interest, so as to allow the zinc finger polypeptide to bind specifically to the sequence of interest.

The term "modifying" as used herein is intended to mean that the sequence is considered modified simply by the binding of the zinc finger polypeptide. It is not intended to suggest that the sequence of nucleotides is changed, although such changes (and others) could ensue following binding of the zinc finger polypeptide to the nucleic acid of interest. Conveniently the nucleic acid sequence is DNA.

Modification of the nucleic acid of interest (in the sense of binding thereto by a zinc finger polypeptide) could be detected in any of a number of methods (e.g. gel mobility shift assays, use of labelled zinc finger polypeptides - labels could include radioactive, fluorescent, enzyme or biotin/streptavidin labels).

Modification of the nucleic acid sequence of interest (and detection thereof) may be all that is required (e.g. in diagnosis of disease). Desirably however, further processing of the sample is performed. Conveniently the zinc finger polypeptide (and nucleic acid

sequences specifically bound thereto) are separated from the rest of the sample. Advantageously the zinc finger polypeptide is bound to a solid phase support, to facilitate such separation. For example, the zinc finger polypeptide may be present in an acrylamide or agarose gel matrix or, more preferably, is immobilised on the surface of a membrane or in the wells of a microtitre place.

Possible uses of suitably designed zinc finger polypeptides are:

- a) Therapy (e.g. targetting to double stranded DNA)
- b) Diagnosis (e.g. detecting mutations in gene sequences:
the present work has shown that "tailor made" zinc finger polypeptides can distinguish DNA sequences differing by one base pair).
- c) DNA purification (the zinc finger polypeptide could be used to purify restriction fragments from solution, or to visualise DNA fragments on a gel [for example, where the polypeptide is linked to an appropriate fusion partner, or is detected by probing with an antibody]).

In addition, zinc finger polypeptides could even be targeted to other nucleic acids such as ss or ds RNA (e.g. self-complementary RNA such as is present in many RNA molecules) or to RNA-DNA hybrids, which would present another possible mechanism of affecting cellular events at the molecular level.

In Example 1 the inventors describe and successfully demonstrate the use of the phage display technique to construct and screen a random zinc finger binding motif library, using a defined oligonucleotide target sequence.

In Example 2 is disclosed the analysis of zinc finger binding motif sequences selected by the screening procedure of Example 1, the DNA-specificity of the motifs being studied by binding to a mini-library of randomised DNA target sequences to reveal a pattern of acceptable bases at each position in the target triplet - a "binding site signature".

In Example 3, the findings of the first two sections are used to select and modify rationally a zinc finger binding polypeptide in order to bind to a particular DNA target with high

affinity: it is convincingly shown that the peptide binds to the target sequence and can modify gene expression in cells cultured *in vitro*.

Example 4 describes the development of an alternative zinc finger binding motif library.

Example 5 describes the design of a zinc finger binding polypeptide which binds to a DNA sequence of special clinical significance.

The invention will now be further described by way of example and with reference to the accompanying drawings, of which:

Figure 1 is a schematic representation of affinity purification of phage particles displaying zinc finger binding motifs fused to phage coat proteins;

Figure 2 shows three amino acid sequences used in the phage display library;

Figure 3 shows the DNA sequences of three oligonucleotides used in the affinity purification of phage display particles;

Figure 4 is a "checker board" of binding site signatures determined for various zinc finger binding motifs;

Figure 5 shows three graphs of fractional saturation against concentration of DNA (nM) for various binding motifs and target DNA triplets;

Figure 6 shows the nucleotide sequence of the fusion between *BCR* and *ABL* sequences in p190 cDNA and the corresponding exon boundaries in the *BCR* and *ABL* genes;

Figure 7 shows the amino acid sequences of various zinc finger binding motifs designed to test for binding to the *BCR/ABL* fusion;

Figure 8 is a graph of peptide binding (as measured by $A_{450-460\text{nm}}$) against DNA

concentration (μM) of target or control DNA sequences;

Figure 9 shows, in the top panel, the result of thin layer chromatography analysis of a chloramphenicol acetyl transferase (CAT) assay, the results of which are represented in the lower panel as a bar chart;

Figure 10 shows photographs of immunofluorescence analysis of various transfected cells (panels A-D);

Figure 11 is a graph showing percentage viability against time for various transfected cells;

Figure 12 shows Northern blot analysis of various transfected cell lines using *ABL*-specific and actin-specific probes;

Figures 13 and 14 illustrate schematically different methods of designing zinc finger binding polypeptides; and

Figure 15 shows the amino acid sequence of zinc fingers in a polypeptide designed to bind to a particular DNA sequence (a ras oncogene).

Example 1

In this example the inventors have used a screening technique to study sequence-specific DNA recognition by zinc finger binding motifs. The example describes how a library of zinc finger binding motifs displayed on the surface of bacteriophage enables selection of fingers capable of binding to given DNA triplets. The amino acid sequences of selected fingers which bind the same triplet were compared to examine how sequence-specific DNA recognition occurs. The results can be rationalised in terms of coded interactions between zinc fingers and DNA, involving base contacts from a few α -helical positions.

An alternative to the rational but biased design of proteins with new specificities, is the isolation of desirable mutants from a large pool. A powerful method of selecting such

proteins is the cloning of peptides (Smith 1985 Science 228, 1315-1317), or protein domains (McCafferty *et al.*, 1990 Nature (London) 348, 552-554; Bass *et al.*, 1990 Proteins 8, 309-314), as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. Phage displaying the peptides of interest can then be affinity purified and amplified for use in further rounds of selection and for DNA sequencing of the cloned gene. The inventors applied this technology to the study of zinc finger-DNA interactions after demonstrating that functional zinc finger proteins can be displayed on the surface of fd phage, and that the engineered phage can be captured on a solid support coated with specific DNA. A phage display library was created comprising variants of the middle finger from the DNA binding domain of Zif268 (a mouse transcription factor containing 3 zinc fingers - Christy *et al.*, 1988). DNA of fixed sequence was used to purify phage from this library over several rounds of selection, returning a number of different but related zinc fingers which bind the given DNA. By comparing similarities in the amino acid sequences of functionally equivalent fingers we deduce the likely mode of interaction of these fingers with DNA. Remarkably, it would appear that many base contacts can occur from three primary positions on the α -helix of the zinc finger, correlating (in hindsight) with the implications of the crystal structure of Zif268 bound to DNA (Pavletich & Pabo 1991). The ability to select or design zinc fingers with desired specificity means that DNA binding proteins containing zinc fingers can now be "made-to-measure".

MATERIALS AND METHODS

Construction and cloning of genes. The gene for the first three fingers (residues 3-101) of Transcription Factor IIIA (TFIIIA) was amplified by PCR from the cDNA clone of TFIIIA using forward and backward primers which contain restriction sites for *NotI* and *SfiI* respectively. The gene for the Zif268 fingers (residues 333-420) was assembled from 8 overlapping synthetic oligonucleotides, giving *SfiI* and *NotI* overhangs. The genes for fingers of the phage library were synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which contained sites for *NotI* and *SfiI* respectively. Backward PCR primers in addition introduced Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these were followed by the residues of the

wild type or library fingers as discussed in the text. Cloning overhangs were produced by digestion with *Sfi*I and *Nor*I where necessary. Fragments were ligated to 1 μ g similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1 (Hoogenboom *et al.*, 1991 Nucleic Acids Res. 19, 4133-4137) in which a section of the *pelB* leader and a restriction site for the enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide (Seq ID No. 1):

5' CTCCTGCAGTTGGACCTGTGCCATGGCCG
GCTGGGCCGCATAGAATGGAACAACATAAGC 3'

which anneals in the region of the polylinker, (L. Jespers, personal communication). Electrocompetent DH5 α cells were transformed with recombinant vector in 200ng aliquots, grown for 1 hour in 2xTY medium with 1% glucose, and plated on TYE containing 15 μ g/ml tetracycline and 1% glucose.

Figure 2 shows the amino acid sequence (Seq ID No. 2) of the three zinc fingers from Zif268 used in the phage display library. The top and bottom rows represent the sequence of the first and third fingers respectively. The middle row represents the sequence of the middle finger. The randomised positions in the α -helix of the middle finger have residues marked 'X'. The amino acid positions are numbered relative to the first helical residue (position 1). For amino acids at positions -1 to +8, excluding the conserved Leu and His, codons are equal mixtures of (G,A,C)NN: T in the first base position is omitted in order to avoid stop codons, but this has the unfortunate effect that the codons for Trp, Phe, Tyr and Cys are not represented. Position +9 is specified by the codon A(G,A)G, allowing either Arg or Lys. Residues of the hydrophobic core are circled, whereas the zinc ligands are written as white letters on black circles. The positions forming the β -sheets and the α -helix of the zinc fingers are marked below the sequence.

Phage selection. Colonies were transferred from plates to 200ml 2xTY/Zn/Tet (2xTY containing 50 μ M Zn(CH₃.COO)₂ and 15 μ g/ml tetracycline) and grown overnight. Phage were purified from the culture supernatant by two rounds of precipitation using 0.2 volumes of 20% PEG/2.5M NaCl containing 50 μ M Zn(CH₃.COO)₂, and resuspended in

zinc finger phage buffer (20mM HEPES pH7.5, 50mM NaCl, 1mM MgCl₂ and 50 μ M Zn(CH₃.COO)₂). Streptavidin-coated paramagnetic beads (Dynal) were washed in zinc finger phage buffer and blocked for 1 hour at room temperature with the same buffer made up to 6% in fat-free dried milk (Marvel). Selection of phage was over three rounds: in the first round, beads (1 mg) were saturated with biotinylated oligonucleotide (~80nM) and then washed prior to phage binding, but in the second and third rounds 1.7nM oligonucleotide and 5 μ g poly dGC (Sigma) were added to the beads with the phage. Binding reactions (1.5ml) for 1 hour at 15°C were in zinc finger phage buffer made up to 2% in fat-free dried milk (Marvel) and 1% in Tween 20, and typically contained 5x10¹¹ phage. Beads were washed 15 times with 1ml of the same buffer. Phage were eluted by shaking in 0.1M triethylamine for 5min and neutralised with an equal volume of 1M Tris pH7.4. Log phase *E. coli* TG1 in 2xTY were infected with eluted phage for 30min at 37°C and plated as described above. Phage titres were determined by plating serial dilutions of the infected bacteria.

The phage selection procedure, based on affinity purification, is illustrated schematically in Figure 1: zinc fingers (A) are expressed on the surface of fd phage(B) as fusions to the minor coat protein (C). The third finger is mainly obscured by the DNA helix. Zinc finger phage are bound to 5'-biotinylated DNA oligonucleotide [D] attached to streptavidin-coated paramagnetic beads [E], and captured using a magnet [F]. (Figure adapted from Dynal AS and also Marks *et al.* (1992 J. Biol. Chem. 267, 16007-16105).

Figure 3 shows sequences (Seq ID No.s 3-8) of DNA oligonucleotides used to purify (i) phage displaying the first three fingers of TFIIIA, (ii) phage displaying the three fingers of Zif268, and (iii) zinc finger phage from the phage display library. The Zif268 consensus operator sequence used in the X-ray crystal structure (Pavletich & Pabo 1991 Science 252, 809-817) is highlighted in (ii), and in (iii) where "X" denotes a base change from the ideal operator in oligonucleotides used to purify phage with new specificities. Biotinylation of one strand is shown by a circled "B".

Sequencing of selected phage. Single colonies of transformants obtained after three rounds of selection as described, were grown overnight in 2xTY/Zn/Tet. Small aliquots

of the cultures were stored in 15% glycerol at -20°C, to be used as an archive. Single-stranded DNA was prepared from phage in the culture supernatant and sequenced using the Sequenase™ 2.0 kit (U.S. Biochemical Corp.).

RESULTS AND DISCUSSION

Phage display of 3-finger DNA-Binding Domains from TFIIIA or Zif268. Prior to the construction of a phage display library, the inventors demonstrated that peptides containing three fully functional zinc fingers could be displayed on the surface of viable fd phage when cloned in the vector Fd-Tet-SN. In preliminary experiments, the inventors cloned as fusions to pIII firstly the three N-terminal fingers from TFIIIA (Ginsberg *et al.*, 1984 Cell 39, 479-489), and secondly the three fingers from Zif268 (Christy *et al.*, 1988), for both of which the DNA binding sites are known. Peptide fused to the minor coat protein was detected in Western blots using an anti-pIII antibody (Stengele *et al.*, 1990 J. Mol. Biol. 212, 143-149). Approximately 10-20% of total pIII in phage preparations was present as fusion protein.

Phage displaying either set of fingers were capable of binding to specific DNA oligonucleotides, indicating that zinc fingers were expressed and correctly folded in both instances. Paramagnetic beads coated with specific oligonucleotide were used as a medium on which to capture DNA-binding phage, and were consistently able to return between 100 and 500-fold more such phage, compared to free beads or beads coated with non-specific DNA. Alternatively, when phage displaying the three fingers of Zif268 were diluted 1:1.7x10³ with Fd-Tet-SN phage not bearing zinc fingers, and the mixture incubated with beads coated with Zif268 operator DNA, one in three of the total phage eluted and transfected into *E. coli* were shown by colony hybridisation to carry the Zif268 gene, indicating an enrichment factor of over 500 for the zinc finger phage. Hence it is clear that zinc fingers displayed on fd phage are capable of preferential binding to DNA sequences with which they can form specific complexes, making possible the enrichment of wanted phage by factors of up to 500 in a single affinity purification step. Therefore, over multiple rounds of selection and amplification, very rare clones capable of sequence-specific DNA binding can be selected-from a large library.

A phage display library of zinc fingers from Zif268. The inventors have made a phage display library of the three fingers of Zif268 in which selected residues in the middle finger are randomised (Figure 2), and have isolated phage bearing zinc fingers with desired specificity using a modified Zif268 operator sequence (Christy & Nathans 1989 Proc. Natl. Acad. Sci. USA 86, 8737-8741) in which the middle DNA triplet is altered to the sequence of interest (Figure 3). In order to be able to study both the primary and secondary putative base recognition positions which are suggested by database analysis (Jacobs 1992 EMBO J. 11, 4507-4517), the inventors have designed the library of the middle finger so that, relative to the first residue in the α -helix (position +1), positions -1 to +8, but excluding the conserved Leu and His, can be any amino acid except Phe, Tyr, Trp and Cys which occur only rarely at those positions (Jacobs 1993 Ph.D. thesis, University of Cambridge). In addition, the inventors have allowed position +9 (which might make an inter-finger contact with Ser at position -2 (Pavletich & Pabo 1991)) to be either Arg or Lys, the two most frequently occurring residues at that position.

The logic of this protocol, based upon the Zif268 crystal structure (Pavletich & Pabo 1991), is that the randomised finger is directed to the central triplet since the overall register of protein-DNA contacts is fixed by its two neighbours. This allows the examination of which amino acids in the randomised finger are the most important in forming specific complexes with DNA of known sequence. Since comprehensive variations are programmed in all the putative contact positions of the α -helix, it is possible to conduct an objective study of the importance of each position in DNA-binding (Jacobs 1992).

The size of the phage display library required, assuming full degeneracy of the 8 variable positions, is $(16^7 \times 2^1) = 5.4 \times 10^8$, but because of practical limitations in the efficiency of transformation with Fd-Tet-SN, the inventors were able to clone only 2.6×10^6 of these. The library used is therefore some two hundred times smaller than the theoretical size necessary to cover all the possible variations of the α -helix. Despite this shortfall, it has been possible to isolate phage which bind with high affinity and specificity to given DNA sequences, demonstrating the remarkable versatility of the zinc finger motif.

Amino acid-base contacts in zinc finger-DNA complexes deduced from phage display selection. Of the 64 base triplets that could possibly form the binding site for variations of finger 2, the inventors have so far used 32 in attempts to isolate zinc finger phage as described. Results from these selections are shown in Table 1, which lists amino acid sequences of the variant α -helical regions from clones of library phage selected after 3 rounds of screening with variants of the Zif268 operator.

Table 1

a	CAG	1	-1123456789	j	ACG	8	-1123456789
		9	RGDHLKDHIK			1	RRDYLMNHIR
			RSDHLTEHIR				RKDVLVSHVR
b	TGA	3	OLAHLSLTHKR	k	ATG	8	RRDYLMNHIR
		1	QSVHLQSHSR			1	RGDALTSHER
		(3)	QKGHLTEHRK			1	RYDALEAHRR
c	GAA	2	OGGNLYRHLR	l	GTA	1	DRSSLTRHTR
		1	NGGNLGRHMK			1	ERTSLSRHIR
		1	ARSNLLRHTR			(1)	GARSLTRHOR
		2	EQSNLYRHOR			(2)	TGGSLARHER
		1	IASNLLRHOR			2	ORASLASHMR
d	GAT	1	DRSNLERHTR	m	TTG	9	RGDALTSHER
		1	NOSNLERHHR			1	RADALMVHKK
		1	QDSNLYRHOR				
		1	NGGNLGRHMK				
		1	NGANLERHRR				
		1	SDGNLORHGR				
		1	SHPNLNRHLK				
		1	TPGNLTRHGR				
e	GAC	4	DRSNLERHTR	n	CCG	5	RQDTLVGHER
		1	DHANLARHTR			1	RQDTLVRHTR
f	GCC	2	DRSSLTRHTR			2	RAADLNRHVR
		7	ERGTLARHEK			1	RKDVLVSHVR
		1	ORRLORHOR			1	RRDYLMNHIR
g	GTC	6	DRSSLTRHTR	o	CCG	1	RSOTLKKHGX
		1	ERTSLSRHIR			3	RGPDLARHGR
h	GCA	1	SAGTLVRHSK			1	AREVLDRHTR
		2	QADTLORHLK			3	REDVLI RHGX
		2	EKATLARHMK			1	RSQLLDRHHK
		1	TGGSLARHER	p	GTC	1	RLOGLRTHLK
i	GCT	1	ROSTLGRHTR			1	RGDALTSHER
		1	EKATLARHMK			1	RADALMVHKK
		1	GAOTLORHLK			1	RYDALEAHRR
		1	ERGTLARHEK			1	RRDYLLNHIR
		1	GRDALARHDK			2	REDVLI RHGX
		1	RGPDLARHGR			1	RSQLLDRHHK
		1	SRDVLRRHNR				

In Table 1, the amino acid sequences, aligned in the one letter code, are listed alongside the DNA oligonucleotides (a to p) used in their purification. The latter are denoted by the sequence of the central DNA triplet in the "bound" strand of the variant Zif268 operator. The amino acid positions are numbered relative to the first helical residue (position 1), and the three primary recognition positions are highlighted. The accompanying numbers indicate the independent occurrences of that clone in the sequenced population (5-10 colonies); where numbers are in parentheses, the clone(s) were detected in the penultimate round of selection but not in the final round. In addition to the DNA triplets shown here, others were also used in attempts to select zinc finger phage from the library, but most selected two clones, one having the α -helical sequence KASNLVSHIR, and the other having the sequence LRHNLETHMR. Those triplets were: ACT, AAA, TTT, CCT, CTT, TTC, AGT, CGA, CAT, AGA, AGC and AAT.

In general the inventors have been unable to select zinc fingers which bind specifically to triplets without a 5' or 3' guanine, all of which return the same limited set of phage after three rounds of selection (see). However for each of the other triplets used to screen the library, a family of zinc finger phage is recovered. In these families is found a sequence bias in the randomised α -helix, which is interpreted as revealing the position and identity of amino acids used to contact the DNA. For instance: the middle fingers from the 8 different clones selected with the triplet GAT (Table 1d) all have Asn at position +3 and Arg at position +6, just as does the first zinc finger of the *Drosophila* protein *tramtrack* in which they are seen making contacts to the same triplet in the cocrystal with specific DNA (Fairall *et al.*, 1993). This indicates that the positional recurrence of a particular amino acid in functionally equivalent fingers is unlikely to be coincidental, but rather because it has a functional role. Thus using data collected from the phage display library (Table 1) it is possible to infer most of the specific amino acid-DNA interactions. Remarkably, most of the results can be rationalised in terms of contacts from the three primary α -helical positions (-1, +3 and +6) identified by X-ray crystallography (Pavletich & Pabo 1991) and database analysis (Jacobs 1992).

As has been pointed out before (Berg 1992 Proc. Natl. Acad. Sci. USA 89, 11109-11110), guanine has a particularly important role in zinc finger-DNA interactions. When present

at the 5' (e.g. Table 1c-i) or 3' (e.g. Table 1m-o) end of a triplet, G selects fingers with Arg at position +6 or -1 of the α -helix respectively. When G is present in the middle position of a triplet (e.g. Table 1b), the preferred amino acid at position +3 is His. Occasionally, G at the 5' end of a triplet selects Ser or Thr at +6 (e.g. Table 1p). Since G can only be specified absolutely by Arg (Seeman *et al.*, 1976 Proc. Nat. Acad. Sci. USA 73, 804-808), this is the most common determinant at -1 and +6. One can expect this type of contact to be a bidentate hydrogen bonding interaction as seen in the crystal structures of Zif268 (Pavletich & Pabo 1991 Science 252, 809-817) and *tramtrack* (Fairall *et al.*, 1993). In these structures, and in almost all of the selected fingers in which Arg recognises G at the 3' end, Asp occurs at position +2 to buttress the long Arg side chain (e.g. Table 1o,p). When position -1 is not Arg, Asp rarely occurs at +2, suggesting that in this case any other contacts it might make with the second DNA strand do not contribute significantly to the stability the protein-DNA complex.

Adenine is also an important determinant of sequence specificity, recognised almost exclusively by Asn or Gln which again are able to make bidentate contacts (Seeman *et al.*, 1976). When A is present at the 3' end of a triplet, Gln is often selected at position -1 of the α -helix, accompanied by small aliphatic residues at +2 (e.g. Table 1b). Adenine in the middle of the triplet strongly selects Asn at +3 (e.g. Table 1c-e), except in the triplet CAG (Table 1a) which selected only two types of finger, both with His at +3 (one being the wild-type Zif268 which contaminated the library during this experiment). The triplets ACG (Table 1j) and ATG (Table 1k), which have A at the 5' end, also returned oligoclonal mixtures of phage, the majority of which were of one clone with Asn at +6.

In theory, cytosine and thymine cannot reliably be discriminated by a hydrogen bonding amino acid side chain in the major groove (Seeman *et al.*, 1976). Nevertheless, C in the 3' position of a triplet shows a marked preference for Asp or Glu at position -1, together with Arg at +1 (e.g. Table 1e-g). Asp is also sometimes selected at +3 and +6 when C is in the middle (e.g. Table 1o) and 5' (e.g. Table 1a) position respectively. Although Asp can accept a hydrogen bond from the amino group of C, one should note that the positive molecular charge of C in the major groove (Hunter 1993 J. Mol. Biol. 230, 1025-1054) will favour an interaction with Asp regardless of hydrogen bonding contacts.

However, C in the middle position most frequently selects Thr (e.g. Table 1i), Val or Leu (e.g. Table 1o) at +3. Similarly, T in the middle position most often selects Ser (e.g. Table 1i), Ala or Val (e.g. Table 1p) at +3. The aliphatic amino acids are unable to make hydrogen bonds but Ala probably has a hydrophobic interaction with the methyl group of C, whereas a longer side chain such as Leu can exclude T and pack against the ring of C. When T is at the 5' end of a triplet, Ser and Thr are selected at +6 (as is occasionally the case for G at the 5' end). Thymine at the 3' end of a triplet selects a variety of polar amino acids at -1 (e.g. Table 1d), and occasionally returns fingers with Ser at +2 (e.g. Table 1a) which could make a contact as seen in the *tramtrack* crystal structure (Fairall *et al.*, 1993).

Limitations of phage display. From Table 1 it can be seen that a consensus or bias usually occurs in two of the three primary positions (-1, +3 and +6) for any family of equivalent fingers, suggesting that in many cases phage selection is by virtue of only two base contacts per finger, as is observed in the Zif268 crystal structure (Pavletich & Pabo 1991). Accordingly, identical finger sequences are often returned by DNA sequences differing by one base in the central triplet. One reason for this is that the phage display selection, being essentially purification by affinity, can yield zinc fingers which bind equally tightly to a number of DNA triplets and so are unable to discriminate. Secondly, since complex formation is governed by the law of mass action, affinity selection can favour those clones whose representation in the library is greatest even though their true affinity for DNA is less than that of other clones less abundant in the library. Phage display selection by affinity is therefore of limited value in distinguishing between permissive and specific interactions beyond those base contacts necessary to stabilise the complex. Thus in the absence of competition from fingers which are able to bind specifically to a given DNA, the tightest non-specific complexes will be selected from the phage library. Consequently, results obtained by phage display selection from a library must be confirmed by specificity assays, particularly when that library is of limited size.

Conclusion. The amino acid sequence biases observed within a family of functionally equivalent zinc fingers indicate that, of the α -helical positions randomised in this study, only three primary (-1, +3 and +6) and one auxiliary (+2) positions are involved in the

recognition of DNA. Moreover, a limited set of amino acids are to be found at those positions, and it is presumed that these make contacts to bases. The indications therefore are that a code can be derived to describe zinc finger-DNA interactions. At this stage however, although sequence homologies are strongly suggestive of amino acid preferences for particular base-pairs, one cannot confidently deduce such rules until the specificity of individual fingers for DNA triplets is confirmed. The inventors therefore defer making a summary table of these preferences until the following example, in which is described how randomised DNA binding sites can be used to this end.

While this work was in progress, a paper by Rebar and Pabo was published (Rebar & Pabo 1994 Science 263, 671-673) in which phage display was also used to select zinc fingers with new DNA-binding specificities. These authors constructed a library in which the first finger of Zif268 is randomised, and screened with tetranucleotides to take into account end effects such as additional contacts from variants of this finger. Only 4 positions (-1, +2, +3 and +6) were randomised, chosen on the basis of the earlier X-ray crystal structures. The results presented above, in which more positions were randomised, to some extent justifies Rebar and Pabo's use of the four random positions without apparent loss of effect, although further selections may reveal that the library is compromised. However, randomising only four positions decreases the theoretical library size so that full degeneracy can be achieved in practice. Nevertheless the inventors found that the results obtained by Rebar and Pabo by screening their complete library with two variant Zif268 operators, are in agreement with their conclusions derived from an incomplete library. On the one hand this again highlights the versatility of zinc fingers but, remarkably, so far both studies have been unable to produce fingers which bind to the sequence CCT. It will be interesting to see whether sequence biases such as we have detected would be revealed, if more selections were performed using Rebar and Pabo's library. In any case, it would be desirable to investigate the effects on selections of using different numbers of randomised positions in more complete libraries than have been used so far.

The original position or context of the randomised finger in the phage display library might bear on the efficacy of selected fingers when incorporated into a new DNA-binding

domain. Selections from a library of the outer fingers of a three finger peptide (Rebar & Pabo, 1994 Science 263, 671-673; Jamieson et al., 1994 Biochemistry 33, 5689-5695) are capable of producing fingers which bind DNA in various different modes, while selections from a library of the middle finger should produce motifs which are more constrained. Accordingly, Rebar and Pabo do not assume that the first finger of Zif268 will always bind a triplet, and screened with a tetranucleotide binding site to allow for different binding modes. Thus motifs selected from libraries of the outer fingers might prove less amenable to the assembly of multifinger proteins, since binding of these fingers could be perturbed on constraining them to a particular binding mode, as would be the case for fingers which had to occupy the middle position of an assembled three-finger protein. In contrast, motifs selected from libraries of the middle finger, having been originally constrained, will presumably be able to preserve their mode of binding even when placed in the outer positions of an assembled DNA-binding domain.

Figure 13 shows different strategies for the design of tailored zinc finger proteins. (A) A three-finger DNA-binding motif is selected *en bloc* from a library of three randomised fingers. (B) A three-finger DNA-binding motif is assembled out of independently selected fingers from a library of one randomised finger (e.g. the middle finger of Zif268). (C) A three-finger DNA-binding motif is assembled out of independently selected fingers from three positionally specified libraries of randomised zinc fingers.

Figure 14 illustrates the strategy of combinatorial assembly followed by *en bloc* selection. Groups of triplet-specific zinc fingers (A) isolated by phage display selection are assembled in random combinations and re-displayed on phage (B). A full-length target site (C) is used to select *en bloc* the most favourable combination of fingers (D).

Example 2

This example describes a new technique to deal efficiently with the selection of a DNA binding site for a given zinc finger (essentially the converse of example 1). This is desirable as a safeguard against spurious selections based on the screening of display libraries. This may be done by screening against libraries of DNA triplet binding sites randomised in two positions but having one base fixed in the third position. The technique is applied here to determine the specificity of fingers previously selected by phage display. The inventors found that some of these fingers are able to specify a unique base in each position of the cognate triplet. This is further illustrated by examples of fingers which can discriminate between closely related triplets as measured by their respective equilibrium dissociation constants. Comparing the amino acid sequences of fingers which specify a particular base in a triplet, we infer that in most instances, sequence specific binding of zinc fingers to DNA can be achieved using a small set of amino acid-base contacts amenable to a code.

One can determine the optimal binding sites of these (and other) proteins, by selection from libraries of randomised DNA. This approach, the principle of which is essentially the converse of zinc finger phage display, would provide an equally informative database from which the same rules can be independently deduced. However until now, the favoured method for binding site determination (involving iterative selection and amplification of target DNA followed by sequencing), has been a laborious process not conveniently applicable to the analysis of a large database (Thiesen & Bach 1990 *Nucleic Acids Res.* 18, 3203-3209; Pollock & Treisman 1990 *Nucleic Acids Res.* 18, 6197-6204).

This example presents a convenient and rapid new method which can reveal the optimal binding site(s) of a DNA binding protein by single step selection from small libraries and use this to check the binding site preferences of those zinc fingers selected previously by phage display. For this application, the inventors have used 12 different mini-libraries of the Zif268 binding site, each one with the central triplet having one position defined with a particular base pair and the other two positions randomised. Each library therefore comprises 16 oligonucleotides and offers a number of potential binding sites to the middle finger, provided that the latter can tolerate the defined base pair. Each zinc finger phage

is screened against all 12 libraries individually immobilised in wells of a microtitre plate, and binding is detected by an enzyme immunoassay. Thus a pattern of acceptable bases at each position is disclosed, which the inventors term a "binding site signature". The information contained in a binding site signature encompasses the repertoire of binding sites recognised by a zinc finger.

The binding site signatures obtained, using zinc finger phage selected as described in example 1, reveal that the selection has yielded some highly sequence-specific zinc finger binding motifs which discriminate at all three positions of a triplet. From measurements of equilibrium dissociation constants it is found that these fingers bind tightly to the triplets indicated in their signatures, and discriminate against closely related sites (usually by at least a factor of ten). The binding site signatures allow progress towards a specificity code for the interactions of zinc fingers with DNA.

MATERIALS AND METHODS

Binding site signatures. Flexible flat-bottomed 96-well microtitre plates (Falcon) were coated overnight at 4°C with streptavidin (0.1mg/ml in 0.1M NaHCO₃ pH8.6, 0.03% NaN₃). Wells were blocked for one hour with PBS/Zn (PBS, 50μM Zn (CH₃.COO)₂) containing 2% fat-free dried milk (Marvel), washed 3 times with PBS/Zn containing 0.1% Tween, and another 3 times with PBS/Zn. The "bound" strand of each oligonucleotide library was made synthetically and the other strand extended from a 5'-biotinylated universal primer using DNA polymerase I (Klenow fragment). Fill-in reactions were added to wells (0.8 pmole DNA library in each) in PBS/Zn for 15 minutes, then washed once with PBS/Zn containing 0.1% Tween, and once again with PBS/Zn. Overnight bacterial cultures each containing a selected zinc finger phage were grown in 2xTY containing 50mM Zn(CH₃.COO)₂ and 15μg/ml tetracycline at 30°C. Culture supernatants containing phage were diluted tenfold by the addition of PBS/Zn containing 2% fat-free dried milk (Marvel), 1% Tween and 20 μg/ml sonicated salmon sperm DNA. Diluted phage solutions (50μl) were applied to wells and binding allowed to proceed for one hour at 20°C. Unbound phage were removed by washing 5 times with PBS/Zn containing 1% Tween, and then 3 times with PBS/Zn. Bound phage were detected as described previously (Griffiths *et al.*, 1994 EMBO J. In press), or using HRP-conjugated anti-M13

IgG (Pharmacia), and quantitated using SOFTmax 2.32 (Molecular Devices Corp).

The results are shown in Figure 4, which gives the binding site signatures of individual zinc finger phage. The figure represents binding of zinc finger phage to randomised DNA immobilised in the wells of microtitre plates. To test each zinc finger phage against each oligonucleotide library (see above), DNA libraries are applied to columns of wells (down the plate), while rows of wells (across the plate) contain equal volumes of a solution of a zinc finger phage. The identity of each library is given as the middle triplet of the "bound" strand of Zif268 operator, where N represents a mixture of all 4 nucleotides. The zinc finger phage is specified by the sequence of the variable region of the middle finger, numbered relative to the first helical residue (position 1), and the three primary recognition positions are highlighted. Bound phage are detected by an enzyme immunoassay. The approximate strength of binding is indicated by a grey scale proportional to the enzyme activity. From the pattern of binding to DNA libraries, called the "signature" of each clone, one or a small number of binding sites can be read off and these are written on the right of the figure.

Determination of apparent equilibrium dissociation constants. Overnight bacterial cultures were grown in 2xTY/Zn/Tet at 30°C. Culture supernatants containing phage were diluted twofold by the addition of PBS/Zn containing 4% fat-free dried milk (Marvel), 2% Tween and 40 µg/ml sonicated salmon sperm DNA. Binding reactions, containing appropriate concentrations of specific 5'-biotinylated DNA and equal volumes of zinc finger phage solution, were allowed to equilibrate for 1h at 20°C. All DNA was captured on streptavidin-coated paramagnetic beads (500µg per well) which were subsequently washed 6 times with PBS/Zn containing 1% Tween and then 3 times with PBS/Zn. Bound phage were detected using HRP-conjugated anti-M13 IgG (Pharmacia) and developed as described (Griffiths *et al.*, 1994). Optical densities were quantitated using SOFTmax 2.32 (Molecular Devices Corp).

The results are shown in Figure 5, which is a series of graphs of fractional saturation against concentration of DNA (nM). The two outer fingers carry the native sequence, as do the the two cognate outer DNA triplets. The sequence of amino acids occupying

helical positions -1 to +9 of the varied finger are shown in each case. The graphs show that the middle finger can discriminate closely related triplets, usually by a factor of ten. The graphs allowed the determination of apparent equilibrium dissociation constants, as below.

Estimations of the K_d are by fitting to the equation $K_d = [\text{DNA}] \cdot [\text{P}] / [\text{DNA} \cdot \text{P}]$, using the KaleidaGraph™ Version 2.0 programme (Abelbeck Software). Owing to the sensitivity of the ELISA used to detect protein-DNA complex, the inventors were able to use zinc finger phage concentrations far below those of the DNA, as is required for accurate calculations of the K_d . The technique used here has the advantage that while the concentration of DNA (variable) must be known accurately, that of the zinc fingers (constant) need not be known (Choo & Klug 1993 Nucleic Acids Res. 21, 3341-3346). This circumvents the problem of calculating the number of zinc finger peptides expressed on the tip of each phage, although since only 10-20% of the gene III protein (pIII) carries such peptides one would expect on average less than one copy per phage. Binding is performed in solution to prevent any effects caused by the avidity (Marks *et al.*, 1992) of phage for DNA immobilised on a surface. Moreover, in this case measurements of K_d by ELISA are made possible since equilibrium is reached in solution prior to capture on the solid phase.

RESULTS AND DISCUSSION

The binding site signature of the second finger of Zif268. The top row of Figure 4 shows the signature of the second finger of wild type Zif268. From the pattern of strong signals indicating binding to oligonucleotide libraries having GNN, TNN, NGN and NNG as the middle triplet, it emerges that the optimal binding site for this finger is T/G,G,G, in accord with the published consensus sequence (Christy & Nathans 1989 Proc. Natl. Acad. Sci. USA 86, 8737-8741). This has implications for the interpretation of the X-ray crystal structure of Zif268 solved in complex with consensus operator having TGG as the middle triplet (Pavletich & Pabo 1991). For instance, His at position +3 of the middle finger was modelled as donating a hydrogen bond to N7 of G, suggesting an equivalent contact to be possible with N7 of A, but from the binding site signature we can see that there is discrimination against A. This implies that the His may prefer to make a

hydrogen bond to O6 of G or a bifurcated hydrogen bond to both O6 and N7, or that a steric clash with the amino group of A may prevent a tight interaction with this base. Thus by considering the stereochemistry of double helical DNA, binding site signatures can give insight into the details of zinc finger-DNA interactions.

Amino acid-base contacts in zinc finger-DNA complexes deduced from binding site signatures. The binding site signatures of other zinc fingers reveal that the phage selections performed in example 1 yielded highly sequence-specific DNA binding proteins. Some of these are able to specify a unique sequence for the middle triplet of a variant Zif268 binding site, and are therefore more specific than is Zif268 itself for its consensus site. Moreover, one can identify the fingers which recognise a particular oligonucleotide library, that is to say a specific base at a defined position, by looking down the columns of Figure 4. By comparing the amino acid sequences of these fingers one can identify any residues which have genuine preferences for particular bases on bound DNA. With a few exceptions, these are as previously predicted on the basis of phage display, and are summarised in Table 2.

Table 2 summarises frequently observed amino acid-base contacts in interactions of selected zinc fingers with DNA. The given contacts comprise a "syllabic" recognition code for appropriate triplets. Cognate amino acids and their positions in the α -helix are entered in a matrix relating each base to each position of a triplet. Auxiliary amino acids from position +2 can enhance or modulate specificity of amino acids at position -1 and these are listed as pairs. Ser or Thr at position +6 permit Asp +2 of the following finger (denoted Asp ++2) to specify both G and T indirectly, and the pairs are listed. The specificity of Ser+3 for T and Thr+3 for C may be interchangeable in rare instances while Val+3 appears to be consistently ambiguous.

Table 2

POSITION IN TRIPLET

	5'	MIDDLE	3'
G	Arg +6 Ser +6/Asp ++2 Thr +6/Asp ++2	His +3	Arg -1/Asp +2
A		Asn +3	Gln -1/Ala +2
T	Ser +6/Asp ++2 Thr +6/Asp ++2	Ala +3 Ser +3 Val +3	Asn -1 Gln -1/Ser +2
C		Asp +3 Leu +3 Thr +3 Val +3	Asp -1

The binding site signatures also reveal an important feature of the phage display library which is important to the interpretation of the selection results. All the fingers in our panel, regardless of the amino acid present at position +6, are able to recognise G or both G and T at the 5' end of a triplet. The probable explanation for this is that the 5' position of the middle triplet is fixed as either G or T by a contact from the invariant Asp at position +2 of finger 3 to the partner of either base on the complementary strand, analogous to those seen in the Zif268 (Pavletich & Pabo 1991 Science 252, 809-817) and *tramtrack* (Fairall *et al.*, 1993) crystal structures (a contact to NH₂ of C or A respectively in the major groove). Therefore Asp at position +2 of finger 3 is dominant over the amino acid present at position +6 of the middle finger, precluding the possibility of recognition of A or C at the 5' position. Future libraries must be designed with this interaction omitted or the position varied. Interestingly, given the framework of the conserved regions of the three fingers, one can identify a rule in the second finger which specifies a frequent interaction with both G and T, viz the occurrence of Ser or Thr at position +6, which may donate a hydrogen bond to either base.

Modulation of base recognition by auxiliary positions. As noted above, position +2 is able to specify the base directly 3' of the 'cognate triplet', and can thus work in conjunction with position +6 of the preceding finger. The binding site signatures, whilst pointing to amino acid-base contacts from the three primary positions, indicate that auxiliary positions can play other parts in base recognition. A clear case in point is Gln at position -1, which is specific for A at the 3' end of a triplet when position +2 is a small non-polar amino acid such as Ala, though specific for T when polar residues such as Ser are at position +2. The strong correlation between Arg at position -1 and Asp at position +2, the basis of which is understood from the X-ray crystal structures of zinc fingers, is another instance of interplay between these two positions. Thus the amino acid at position +2 is able to modulate or enhance the specificity of the amino acid at other positions.

At position +3, a different type of modulation is seen in the case of Thr and Val which most often prefer C in the middle position of a triplet, but in some zinc fingers are able to recognise both C and T. This ambiguity occurs possibly as a result of different

hydrophobic interactions involving the methyl groups of these residues, and here a flexibility in the inclination of the finger rather than an effect from another position *per se* may be the cause of ambiguous reading.

Quantitative measurements of dissociation constants. The binding site signature of a zinc finger reveals its differential base preferences at a given concentration of DNA. As the concentration of DNA is altered, one can expect the binding site signature of any clone to change, being more distinctive at low [DNA], and becoming less so at higher [DNA] as the K_d of less favourable sites is approached and further bases become acceptable at each position of the triplet. Furthermore, because two base positions are randomly occupied in any one library of oligonucleotides, binding site signatures are not formally able to exclude the possibility of context dependence for some interactions. Therefore to supplement binding site signatures, which are essentially comparative, quantitative determinations of the equilibrium dissociation constant of each phage for different DNA binding sites are required. After phage display selection and binding site signatures, these are the third and definitive stage in assessing the specificity of zinc fingers.

Examples of such studies presented in Figure 5 reveal that zinc finger phages bind the operators indicated in their binding site signatures with K_d s in the range of 10^{-8} - 10^{-9} M. and can discriminate against closely related binding sites by factors greater than an order of magnitude. Indeed Figure 5 shows such differences in affinity for binding sites which differ in only one out of nine base pairs. Since the zinc fingers in our panel were selected from a library by non-competitive affinity purification, there is the possibility that fingers which are even more discriminatory can be isolated using a competitive selection process.

Measurements of dissociation constants allow different triplets to be ranked in order of preference according to the strength of binding. The examples here indicate that the contacts from either position -1 or +3 can contribute to discrimination. Also, the ambiguity in certain binding site signatures referred to above can be shown to have a basis in the equal affinity of certain figures for closely related triplets. This is demonstrated by the K_d s of the finger containing the amino acid sequence RGDALTSHER for the triple TTG and GTG.

A code for zinc finger-DNA recognition. One would expect that the versatility of the zinc finger motif will have allowed evolution to develop various modes of binding to DNA (and even to RNA), which will be too diverse to fall under the scope of a single code. However, although a code may not apply to all zinc finger-DNA interactions, there is now convincing evidence that a code applies to a substantial subset. This code will fall short of being able to predict unfailingly the DNA binding site preference of any given zinc finger from its amino acid sequence, but may yet be sufficiently comprehensive to allow the design of zinc fingers with specificity for a given DNA sequence.

Using the selection methods of phage display (as described above) and of binding site signatures it is found that in the case of Zif268-like zinc fingers, DNA recognition involves four fixed principal (three primary and one auxiliary) positions on the α -helix, from where a limited and specific set of amino acid-base contacts result in recognition of a variety of DNA triplets. In other words, a code can describe the interactions of zinc fingers with DNA. Towards this code, one can propose amino acid-base contacts for almost all the entries in a matrix relating each base to each position of a triplet (Table 2). Where there is overlap, the results presented here complement those of Desjarlais and Berg who have derived similar rules by altering zinc finger specificity using database-guided mutagenesis (Desjarlais & Berg 1992 Proc Natl. Acad. Sci. USA 89, 7345-7349; Desjarlais & Berg 1993 Proc. Natl. Acad. Sci. USA 90, 2256-2260).

Combinatorial use of the coded contacts. The individual base contacts listed in Table 2, though part of a code, may not always result in sequence specific binding to the expected base triplet when used in any combination. In the first instance one must be aware of the possibility that zinc fingers may not be able to recognise certain combinations of bases in some triplets by use of this code, or even at all. Otherwise, the majority of inconsistencies may be accounted for by considering variations in the inclination of the trident reading head of a zinc finger with respect to the triplet with which it is interacting. It appears that the identity of an amino acid at any one α -helical position is attuned to the identity of the residues at the other two positions to allow three base contacts to occur simultaneously. Therefore, for example, in order that Ala may pick out T in the triplet GTG, Arg must not be used to recognise G from position +6, since this would distance

the former too far from the DNA (see for example the finger containing the amino acid sequence RGDALTSHER). Secondly, since the pitch of the α -helix is 3.6 amino acids per turn, positions -1, +3 and +6 are not an integral number of turns apart, so that position +3 is nearer to the DNA than are -1 or +6. Hence, for example, short amino acids such as His and Asn, rather than the longer Arg and Gln, are used for the recognition of purines in the middle position of a triplet.

As a consequence of these distance effects one might say that the code is not really "alphabetic" (always identical amino acid:base contact) but rather "syllabic" (use of a small repertoire of amino acid:base contacts). An alphabetic code would involve only four rules, but syllabicity adds an additional level of complexity, since systematic combinations of rules comprise the code. Nevertheless, the recognition of each triplet is still best described by a code of syllables, rather than a catalogue of "logograms" (idiosyncratic amino acid:base contact depending on triplet).

Conclusions. The "syllabic" code of interactions with DNA is made possible by the versatile framework of the zinc finger: this allows an adaptability at the interface with DNA by slight changes of orientation, which in turn maintains a stoichiometry of one coplanar amino acid per base-pair in many different complexes. Given this mode of interaction between amino acids and bases it is to be expected that recognition of G and A by Arg and Asn/Gln respectively are important features of the code; but remarkably other interactions can be more discriminatory than was anticipated (Seeman *et al.*, 1976). Conversely, it is clear that degeneracy can be programmed in the zinc fingers in varying degrees allowing for intricate interactions with different regulatory DNA sequences (Harrison & Travers, 1990; Christy & Nathans, 1989). One can see how this principle makes possible the regulation of differential gene expression by a limited set of transcription factors.

As already noted above, the versatility of the finger motif will likely allow other modes of binding to DNA. Similarly, one must take into account the malleability of nucleic acids such as is observed in Fairall *et al.*, where a deformation of the double helix at a flexible base step allows a direct contact from Ser at position +2 of finger 1 to a T at the 3'

position of the cognate triplet. Even in our selections there are instances of fingers whose binding mode is obscure, and may require structural analyses for clarification. Thus, water may be seen to play an important role, for example where short side chains such as Asp, Asn or Ser interact with bases from position -1 (Qian *et al.*, 1993 J. Am. Chem. Soc. 115, 1189-1190; Shakked *et al.*, 1994 Nature (London) 368, 469-478).

Eventually, it might be possible to develop a number of codes describing zinc finger binding to DNA, which could predict the binding site preferences of some zinc fingers from their amino acid sequence. The functional amino acids selected at positions -1, +3 and to an extent +6 in this study, are very frequently observed at the same positions in naturally occurring fingers (e.g. see Fig. 4. of Desjarlais and Berg 1992 Proteins 12, 101-104) supporting the existence of coded contacts from these three positions. However, the lack of definitive predictive methods is not a serious practical limitation as current laboratory techniques (here and in Thiesen & Bach 1990 and Pollock & Treisman 1990) will allow the identification of binding sites for a given DNA-binding protein. Rather, one can apply phage selection and a knowledge of the recognition rules to the converse problem, namely the design of proteins to bind predetermined DNA sites.

Prospects for the design of DNA-binding proteins. The ability to manipulate the sequence specificity of zinc fingers implies that we are on the eve of designing DNA-binding proteins with desired specificity for applications in medicine and research (Desjarlais & Berg, 1993; Rebar & Pabo, 1994). This is possible because, by contrast to all other DNA-binding motifs, we can avail ourselves of the modular nature of the zinc finger, since DNA sites can be recognised by appropriate combinations of independently acting fingers linked in tandem.

The coded interactions of zinc fingers with DNA can be used to model the specificity of individual zinc fingers *de novo*, or more likely in conjunction with phage display selection of suitable candidates. In this way, according to requirements, one could modulate the affinity for a given binding site, or even engineer an appropriate degree of indiscrimination at particular base positions. Moreover, the additive effect of multiply repeated domains offers the opportunity to bind specifically and tightly to extended, and

hence very rare, genomic loci. Thus zinc finger proteins might well be a good alternative to the use of antisense nucleic acids in suppressing or modifying the action of a given gene, whether normal or mutant. To this end, extra functions could be introduced to these DNA binding domains by appending suitable natural or synthetic effectors.

Example 3

From the evidence presented in the preceding examples, the inventors propose that specific DNA-binding proteins comprising zinc fingers can be "made to measure". To demonstrate their potential the inventors have created a three finger polypeptide able to bind site-specifically to a unique 9bp region of a *BCR-ABL* fusion oncogene and to discriminate it from the parent genomic sequences (Kurzrock *et al.*, 1988 N. Engl. J. Med. 319, 990-998). Using transformed cells in culture as a model, it is shown that binding to the target oncogene in chromosomal DNA is possible, resulting in blockage of transcription. Consequently, murine cells made growth factor-independent by the action of the oncogene (Daley *et al.*, 1988 Proc. Natl. Acad. Sci. U.S.A. 85, 9312-9316) are found to revert to factor dependence on transient transfection with a vector expressing the designed zinc finger polypeptide.

DNA-binding proteins designed to recognise specific DNA sequences could be incorporated in chimeric transcription factors, recombinases, nucleases etc. for a wide range of applications. The inventors have shown that zinc finger mini-domains can discriminate between closely related DNA triplets, and have proposed that they can be linked together to form domains for the specific recognition of longer DNA sequences. One interesting possibility for the use of such protein domains is to target selectively genetic differences in pathogens or transformed cells. Here one such application is described.

There exist a set of human leukaemias in which a reciprocal chromosomal translocation $t(9;22)(q34;q11)$ result in a truncated chromosome 22, the Philadelphia chromosome (Ph1)5, encoding at the breakpoint a fusion of sequences from the *c-ABL* protooncogene (Bartram *et al.*, 1983 Nature 306, 277-280) and the *BCR* gene (Groffen *et al.*, 1984 Cell 36, 93-99). In chronic myelogenous leukaemia (CML), the breakpoints usually occur in

the first intron of the c-*ABL* gene and in the breakpoint cluster region of the *BCR* gene (Shtivelman *et al.*, 1985 Nature 315, 550-554), and give rise to a p210^{*BCR-ABL*} gene product (Konopka *et al.*, 1984 Cell 37, 1035-1042). Alternatively, in acute lymphoblastic leukaemia (ALL), the breakpoints usually occur in the first introns of both *BCR* and c-*ABL* (Hermans *et al.*, 1987 Cell 51, 33-40), and result in a p190^{*BCR-ABL*} gene product (Figure 6) (Kurzrock *et al.*, 1987 Nature 325, 631-635).

Figure 6 shows the nucleotide sequences (Seq ID No.s 9-11) of the fusion point between *BCR* and *ABL* sequences in p190 cDNA, and of the corresponding exon boundaries in the *BCR* and c-*ABL* genes. Exon sequences are written in capital letters while introns are given in lowercase. Line 1 shows p190^{*BCR-ABL*} cDNA; line 2 the *BCR* genomic sequence at junction of exon 1 and intron 1; and line 3 the *ABL* genomic sequence at junction of intron 1 and exon 2 (Hermans *et al* 1987). The 9bp sequence in the p190^{*BCR-ABL*} cDNA used as a target is underlined, as are the homologous sequences in genomic *BCR* and c-*ABL*.

Facsimiles of these rearranged genes act as dominant transforming oncogenes in cell culture (Daley *et al.*, 1988) and transgenic mice (Heisterkamp *et al.*, 1990 Nature 344, 251-253). Like their genomic counterparts, the cDNAs bear a unique nucleotide sequence at the fusion point of the *BCR* and c-*ABL* genes, which can be recognised at the DNA level by a site-specific DNA-binding protein. The present inventors have designed such a protein to recognise the unique fusion site in the p190^{*BCR-ABL*} c-DNA. This fusion is obviously distinct from the breakpoints in the spontaneous genomic translocations, which are thought to be variable among patients. Although the design of such peptides has implications for cancer research, the primary aim here is to prove the principle of protein design, and to assess the feasibility of *in vivo* binding to chromosomal DNA in available model systems.

A nine base-pair target sequence (GCA, GAA, GCC) for a three zinc finger peptide was chosen which spanned the fusion point of the p190^{*BCR-ABL*} cDNA (Hermans *et al.*, 1987). The three triplets forming this binding site were each used to screen a zinc finger phage library over three rounds as described above in example 1. The selected fingers were then analysed by binding site signatures to reveal their preferred triplet, and mutations to

improve specificity were made to the finger selected for binding to GCA. A phage display mini-library of putative *BCR-ABL*-binding three-finger proteins was cloned in fd phage, comprising six possible combinations of the six selected or designed fingers (1A, 1B; 2A; 3A, 3B and 3C) linked in the appropriate order. These fingers are illustrated in Figure 7 (Seq ID No.s 12-17). In Figure 7 regions of secondary structure are underlined below the list, while residue positions are given above, relative to the first position of the α -helix (position 1). Zinc finger phages were selected from a library of 2.6×10^6 variants, using three DNA binding sites each containing one of the triplets GCC, GAA or GCA. Binding site signatures (example 2) indicate that fingers 1A and 1B specify the triplet GCC, finger 2A specifies GAA, while the fingers selected using the triplet GCA all prefer binding to GCT. Amongst the latter is finger 3A, the specificity of which we believed, on the basis of recognition rules, could be changed by a point mutation. Finger 3B, based on the selected finger 3A, but in which Gln at helical position +2 was altered to Ala should be specific for GCA. Finger 3C is an alternative version of finger 3A, in which the recognition of C is mediated by Asp+3 rather than by Thr+3.

The mini library was screened once with an oligonucleotide containing the 9 base-pair *BCR-ABL* target sequence to select for tight binding clones over weak binders and background vector phage. Because the library was small, the inventors did not include competitor DNA sequences for homologous regions of the genomic *BCR* and *c-ABL* genes but instead checked the selected clones for their ability to discriminate. It was found that although all the selected clones were able to bind the *BCR-ABL* target sequence and to discriminate between this and the genomic-*BCR* sequence, only a subset could discriminate against the *c-ABL* sequence which, at the junction between intron 1 and exon 2, has an 8/9 base-pair homology to the *BCR-ABL* target sequence (Hermans *et al.*, 1987). Sequencing of the discriminating clones revealed two types of selected peptide, one with the composition 1A-2A-3B and the other with 1B-2A-3B. Thus both peptides carried the third finger (3B) which was specifically designed against the triplet GCA but peptide 1A-2A-3B was able to bind to the *BCR-ABL* target sequence with higher affinity than was peptide 1B-2A-3B.

The peptide 1A-2A-3B, henceforth referred to as the anti-*BCR-ABL* peptide, was used in

further experiments. The anti-BCR-ABL peptide has an apparent equilibrium dissociation constant (K_d) of $6.2 \pm 0.4 \times 10^{-7} \text{M}$ for the p190^{BCR-ABL} cDNA sequence *in vitro*, and discriminates against the similar sequences found in genomic *BCR* and *c-ABL* DNA, by factors greater than an order of magnitude (Figure 8). Referring to Figure 8, (which illustrates discrimination in the binding of the anti-BCR-ABL peptide to its p190^{BCR-ABL} target site and to like regions of genomic *BCR* and *c-ABL*), the graph shows binding (measured as an $A_{450-650}$) at various [DNA]. Binding reactions and complex detection by enzyme immunoassay were performed as described previously, and a full curve analysis was used in calculations of the K_d (Choo & Klug 1993). The DNA used were oligonucleotides spanning 9bp either side of the fusion point in the cDNA or the exon boundaries. The anti-BCR-ABL peptide binds to its intended target site with a $K_d = 6.2 \pm 0.4 \times 10^{-7} \text{M}$, and is able to discriminate against genomic *BCR* and *c-ABL* sequences, though the latter differs by only one base pair in the bound 9bp region.

The measured dissociation constant is higher than that of three-finger peptides from naturally occurring proteins such as Spl (Kadonga *et al.*, 1987 Cell 51, 1079-1090) or Zif268 (Christy *et al.*, 1988), which have K_d s in the range of 10^{-9}M , but rather is comparable to that of the two fingers from the *tramtrack* (*ttk*) protein (Fairall *et al.*, 1992). However, the affinity of the anti-BCR-ABL peptide could be refined, if desired, by site-directed mutations or by "affinity maturation" of a phage display library (Hawkins *et al.*, 1992 J. Mol. Biol. 226, 889-896).

Having established DNA discrimination *in vitro*, the inventors wished to test whether the anti-BCR-ABL peptide was capable of site-specific DNA-binding *in vivo*. The peptide was fused to the VP16 activation domain from herpes simplex virus (Fields 1993 Methods 5, 116-124) and used in transient transfection assays (Figure 9) to drive production of a CAT (chloramphenicol acetyl transferase) reporter gene from a binding site upstream of the TATA box (Gorman *et al.*, Mol. Cell. Biol. 2, 1044-1051). In detail, the experiment was performed thus: reporter plasmids pMCAT6BA, pMCAT6A, and pMCAT6B, were constructed by inserting 6 copies of the p190^{BCR-ABL} target site (CGCAGAAGCC), the *c-ABL* second exon-intron junction sequence (TCCAGAAGCC), or the *BCR* first exon-intron junction sequence (CGCAGGTGAG) respectively, into pMCAT3 (Luscher *et al.*, 1989 Genes Dev. 3, 1507-1517). The anti-BCR-ABL/VP16 expression vector was

generated by inserting the in-frame fusion between the activation domain of herpes simplex virus VP16 (Fields 1993) and the Zn finger peptide in the pEF-BOS vector (Mizushima & Shigezaku 1990 Nucl. Acids Res. 18, 5322). C3H10T1/2 cells were transiently co-transfected with 10 μ g of reporter plasmid and 10 μ g of expression vector. RSVL (dWet *et al.*, 1987 Mol. Cell Biol. 7, 725-737), which contains the Rous sarcoma virus long terminal repeat linked to luciferase, was used as an internal control to normalise for differences in transfection efficiency. Cells were transfected by the calcium phosphate precipitation method and CAT assays performed as described (Sanchez-Garcia *et al.*, 1999 EMBO J. 12, 4243-4250). Plasmid pGSEC, which has five consensus 17-mer GAL4-binding sites upstream from the minimal promoter of the adenovirus Elb TAT box, and pMIVP16 vector, which encodes an in-frame fusion between the DNA-binding domain of GAL4 and the activation domain of herpes simplex virus VP16, were used as a positive control (Sadowski *et al.*, 1992 Gene 118, 137-141). The results are shown in Figure 9.

Referring to Figure 9, C3H10T1/2 cells were transiently cotransfected with a CAT reporter plasmid and an anti-BCR-ABL/VP16 expression vector (pZNL1A). The top panel of the figure shows the results of thin layer chromatography of samples from different transfections, in which the fold induction of CAT activity relative to a sample where reporter alone was transfected (panel 1) is plotted on a histogram below.

A specific (thirty-fold) increase in CAT activity was observed in cells cotransfected with reporter plasmid bearing copies of the p190^{BCR-ABL} cDNA target site, compared to a barely detectable increase in cells cotransfected with reporter plasmid bearing copies of either the BCR or c-ABL semihomologous sequences, indicating *in vivo* binding. The particular constructs used in different transfections are noted below the histogram.

The selective stimulation of transcription indicates convincingly that highly site-specific DNA-binding can occur *in vivo*. However, while transient transfections assay binding of plasmid DNA, the true target site for this and most other DNA-binding proteins is genomic DNA. This might well present significant problems, not least since this DNA is physically separated from the cytosol by the nuclear membrane, but also since it may

be packaged within chromatin.

To study whether genomic targeting is possible, a construct was made in which the anti-BCR-ABL peptide was flanked at the N-terminus with the nuclear localisation signal from the large T antigen of SV40 virus (Kalderon *et al.*, 1984 Cell 499-509), and at the C-terminus with an 11 amino acid c-myc epitope tag recognisable by the 9E10 antibody (Evan *et al.*, 1985 Mol. Cell. Biol. 5, 3610-3616). This construct was used to transiently transfect the IL-3-dependent murine cell line Ba/F3 (Palacios & Steinmetz 1985 Cell 41 727-734), or alternatively Ba/F3+p190 and Ba/F3+p210 cell lines previously made IL-3-independent by integrated plasmid constructs expressing either p190^{BCR-ABL} or p210^{BCR-ABL}, respectively. Staining of the cells with the 9E10 antibody followed by secondary fluorescent conjugate showed efficient nuclear localisation in those cells transfected with the anti-BCR-ABL peptide.

The experimental details were as follows: the anti-BCR-ABL expression vector was generated in the pEF-BOS vector (Mizushima & Shigezaku 1990), including an 11 amino acid c-myc epitope tag (EQKLISEEDLN) at the carboxy-terminal end, recognizable by the 9E10 antibody (Evan *et al.*, 1985) and the nuclear localization signal PKKKRKV of the large T antigen of SV40 virus (Kalderon *et al.*, 1984) at the amino-terminal end. Three glycine residues were introduced downstream of the nuclear localization signal as a spacer to ensure exposure of the nuclear leader from the folded molecule. Ba/F3 cells were transfected with 25 µg of the anti-BCR-ABL expression construct tagged with the 9E10 c-myc epitope as described (Sanchez-Garcia & Rabbitts 1994 Proc. Natl. Acad. Sci U.S.A. in press) and protein production analyzed 48 h later by immunofluorescence-labelling as follows. Cells were fixed in 4% (w/v) paraformaldehyde for 15 min, washed in phosphate-buffered saline (PBS), and permeabilized in methanol for 2 min. After blocking in 10% fetal calf serum in PBS for 30 min, the mouse 9E10 antibody was added. After a 30 min incubation at room temperature a fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG (SIGMA) was added and incubated for a further 30 min. Fluorescent cells were visualized using a confocal scanning microscope (magnification, 200X). The results are shown in Figure 10.

Figure 10 (immunofluorescence of Ba/F3+p190 and Ba/F3+p210 cells transiently transfected with the anti-bcr-abl expression vector and stained with the 9E10 antibody), the image shows expression and nuclear localisation of the anti-BCR-ABL peptide (panels A, C, and D). In addition, transfected Ba/F3+p190 cells show chromatin condensation and nuclear fragmentation into small apoptotic bodies (panels B, and C), but not either untransfected Ba/F3+p190 cells (panel A) or transfected Ba/F3+p210 cells (panel D).

The efficiency of transient transfection, measured as the proportion of immunofluorescent cells in the population, was 15-20%. When IL-3 is withdrawn from tissue culture, a corresponding proportion of Ba/F3+p190 cells are found to have reverted to factor dependence and die, while Ba/F3+p210 cells are unaffected. The experimental details were as follows: cell lines Ba/F3, Ba/F3+p190 and Ba/F3+p210 were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. In the case of Ba/F3 cell line 10% WEHI-3B-conditioned medium was included as a source of IL-3. After the transfection with the anti-BCR-ABL expression vector, cells (5×10^5 /ml) were washed twice in serum-free medium and cultured in DMEM medium with 10% fetal bovine serum without WEHI-3B-conditioned medium. Percentage viability was determined by trypan blue exclusion. Data are expressed as means of triplicate cultures. The results are shown in graphical form in Figure 11.

Immunofluorescence microscopy of transfected Ba/F3+p190 cells in the absence of IL-3 shows chromatin condensation and nuclear fragmentation into small apoptotic bodies, while the nuclei of Ba/F3+p210 cells remain intact (Figure 10). Northern blots of total cytoplasmic RNA from Ba/F3+p190 cells transiently transfected with the anti-BCR-ABL peptide revealed reduced levels of p190^{BCR-ABL} mRNA relative to untransfected cells. By contrast, similarly transfected Ba/F3+p210 cells showed no decrease in the levels of p210^{BCR-ABL} mRNA (Figure 12). The blots were performed as follows: 10 μ g of total cytoplasmic RNA, from the cells indicated, was glyoxylated and fractionated in 1.4% agarose gels in 10mM NaPO₄ buffer, pH 7.0. After electrophoresis the gel was blotted onto Hybond-N (Amersham), UV-cross linked and hybridized to an ³²P-labelled c-ABL probe. Autoradiography was for 14h at -70°C. Loading was monitored by reprobing the filters with a mouse β -actin cDNA.

Referring to Figure 12, (Northern filter hybridisation analysis of Ba/F3+p190 and Ba/F3+p210 cell lines transfected with the anti-BCR-ABL expression vector), lane 1 is from untransfected Ba/F3+p190 cell line; lanes 2, and 3 are from Ba/F3+p190 cell line transfected with the anti-BCR-ABL expression vector; lane 4 is from untransfected Ba/F3+p210 cell line; lanes 5 and 6 are from Ba/F3+p210 cell line transfected with the anti-BCR-ABL expression vector. When transfected with the anti-BCR-ABL expression vector, a specific downregulation of p190^{BCR-ABL} mRNA is seen in Ba/F3+p190 cells, while expression of p210^{BCR-ABL} is unaffected in Ba/F3+p210 cells.

In summary, the inventors have demonstrated that a DNA-binding protein designed to recognise a specific DNA sequence *in vitro*, is active *in vivo* where, directed to the nucleus by an appended localisation signal, it can bind its target sequence in chromosomal DNA. This is found on otherwise actively transcribing DNA, so presumably binding of the peptide blocks the path of the polymerase, causing stalling or abortion. The use of a specific polypeptide in this case to target intragenic sequences is reminiscent of antisense oligonucleotide- or ribozyme- based approaches to inhibiting the expression of selected genes (Stein & Cheng 1993 Science 261, 1004-1012). Like antisense oligonucleotides, zinc finger DNA-binding proteins can be tailored against genes altered by chromosomal translocations, or point mutations, as well as to regulatory sequences within genes. Also, like oligonucleotides which can be designed to repress transcription by triple helix formation in homopurine-homopyrimidine promoters (Cooney *et al.*, 1988 Science 245, 725-730) DNA-binding proteins can bind to various unique regions outside genes, but in contrast they can direct gene expression by both up- or down- regulating, the initiation of transcription when fused to activation (Seipel *et al.*, 1992 EMBO J. 11, 4961-4968) or repression domains (Herschbach *et al.*, 1994 Nature 370, 309-311). In any case, by acting directly on any DNA, and by allowing fusion to a variety of protein effectors, tailored site-specific DNA-binding proteins have the potential to control gene expression, and indeed to manipulate the genetic material itself, in medicine and research.

Example 4

The phage display zinc finger library described in the preceding examples could be

considered sub-optimal in a number of ways:-

- i) the library was much smaller than the theoretical maximum size;
- ii) the flanking fingers both recognised GCG triplets (in certain cases creating nearly symmetrical binding sites for the three zinc fingers, which enables the peptide to bind to the 'bottom' strand of DNA, thus evading the register of interactions we wished to set);
- iii) Asp+2 of finger three ("Asp++2") was dominant over the interactions of finger two (position +6) with the 5' base of the middle triplet;
- iv) not all amino acids were represented in the randomised positions.

In order to overcome these problems a new three-finger library was created in which:

- a) the middle finger is fully randomised in only four positions (-1, +2, +3 and +6) so that the library size is smaller and all codons are represented. The library was cloned in the pCANTAB5E phagemid vector from Pharmacia, which allows higher transformation frequencies than the phage.
- b) the first and third fingers recognise the triplets GAC and GCA, respectively, making for a highly asymmetric binding site. Recognition of the 3'A in the latter triplet by finger three is mediated by Gln-1/Ala+2, the significance of which is that the short Ala+2 should not make contacts to the DNA (in particular with the 5' base of the middle triplet), thus alleviating the problem noted at (iii) above.

Example 5

The human ras gene is susceptible to a number of different mutations, which can convert it into an oncogene. A ras oncogene is found in a large number of human cancers. One particular mutation is known as the G12V mutation (i.e. the polypeptide encoded by the mutant gene contains a substitution from glycine to valine). Because ras oncogenes are

so common in human cancers, they are extremely significant targets for potential therapeutic methods.

A three finger protein has been designed which can recognise the G12V mutant of ras. The protein was produced using rational design based on the known specificity rules. In outline, a zinc finger framework (from one of the fingers selected to bind GCC) was modified by point mutations in position +3 to yield fingers recognising two additional different triplets. The finger recognising GCC and the two derivatives were cloned in pCANTAB5E and expressed on the surface of phage.

Originally, the G12V-binding peptide "r-BP" was to be selected from a small library of related proteins. The reason a library was to be used is that while it was clear to us what 8/9 of the amino acid:base contacts should be, it was not clear whether the middle C of the GCC triplet should be recognised by +3 Asp, or Glu, or Ser, or Thr (see Table 2 above). Thus a three-finger peptide gene was assembled from 8 overlapping synthetic oligonucleotides which were annealed and ligated according to standard procedures and the ~300bp product purified from a 2% agarose gel. The gene for finger 1 contained a partial codon randomisation at position +3 which allowed for inclusion of each of the above amino acids (D, E, S & T) and also certain other residues which were in fact not predicted to be desirable (e.g. Asn). The synthetic oligonucleotides were designed to have *Sfi*I and *Nor*I overhangs when annealed. The ~300bp fragment was ligated into *Sfi*I/*Nor*I-cut FdSN vector and the ligation mixture was electroporated into DH5 α cells. Phage were produced from these as previously described and a selection step carried out using the G12V sequence (also as described) to eliminate phage without insert and those phage of the library which bound poorly.

Following selection, a number of separate clones were isolated and phage produced from these were screened by ELISA for binding to the G12V ras sequence and discrimination against the wild-type ras sequence. A number of clones were able to do this, and sequencing of phage DNA later revealed that these fell into two categories, one of which had the amino acid Asn at the +3 randomised position, and another which had two other undesirable mutations.

The appearance of Asn at position +3 is unexpected and most probably due to the fact that proteins with a cytosine-specific residue at position +3 bind to some *E. coli* DNA sequence so tightly that they are lethal. Thus phage display selection is not always guaranteed to produce the tightest-binding clone, since passage through bacteria is essential to the technique, and the selected proteins may be those which do not bind to the genome of this host if such binding is deleterious.

K_d measurements show that the clone with Asn+3 nevertheless binds the mutant G12V sequence with a K_d in the nM range and discriminates against the wild-type ras sequence. However it was predicted that Asn+3 should specify an adenine residue at the middle position, whereas the polypeptide we wished to make should specify a cytosine for optimal binding.

Thus we assembled a three-finger peptide with a Ser at position +3 of Finger 1 (as shown in Figure 15), again for using synthetic oligos. This time the gene was ligated to pCANTAB5E phagemid. Transformants were isolated in the *E. coli* ABLE-C strain (from Stratagene) and grown at 30°C, which strain under these conditions reduces the copy number of plasmids so as to make their toxic products less abundant in the cells.

The amino acid sequence (Seq ID No. 18) of the fingers is shown in Figure 15. The numbers refer to the α -helical amino acid residues. The fingers (F1, F2 & F3) bind to the G12V mutant nucleotide sequence: 5' GAC GGC GCC 3'

F3 F2 F1

The bold A shows the single point mutation by which the G12V sequence differs from the wild type sequence.

Assay of the protein in eukaryotes (e.g. to drive CAT reporter production) requires the use of a weak promoter. When expression of the anti-RAS (G12V) protein is strong, the peptide presumably binds to the wild-type ras allele (which is required) leading to cell death. For this reason, a regulatable promoter (e.g. for tetracycline) will be used to deliver the protein in therapeutic applications, so that the intracellular concentration of the

protein exceeds the K_d for the G12V point mutated gene but not the K_d for the wild-type allele. Since the G12V mutation is a naturally occurring genomic mutation (not only a cDNA mutation as was the p190 bcr-abl) human cell lines and other animal models can be used in research.

In addition to repressing the expression of the gene, the protein can be used to diagnose the precise point mutation present in the genomic DNA, or more likely in PCR amplified genomic DNA, without sequencing. It should therefore be possible, without further inventive activity, to design diagnostic kits for detecting (e.g. point) mutations on DNA. ELISA-based methods should prove particularly suitable.

It is hoped to fuse the zinc finger binding polypeptide to an scFv fragment which binds to the human transferrin receptor, which should enhance delivery to and uptake by human cells. The transferrin receptor is thought particularly useful but, in theory, any receptor molecule (preferably of high affinity) expressed on the surface of a human target cell could act as a suitable ligand, either for a specific immunoglobulin or fragment, or for the receptor's natural ligand fused or coupled with the zinc finger polypeptide.

SEQUENCE LISTING

1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Medical Research Council
 (B) STREET: 20 Park Crescent
 (C) CITY: London
 (E) COUNTRY: United Kingdom
 (F) POSTAL CODE (ZIP): W1N 4AL

(ii) TITLE OF INVENTION: Improvements in or Relating to Binding
 Proteins for Recognition of DNA

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0. Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTCTGTCAGT TGGACCTGTG CCATGGCCGG CTGGGCCGCA TAGAATGGAA CAACTAAAGC

60

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Glu Glu Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
 1 5 10 15

Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
 20 25 30

Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Xaa
 35 40 45
 Xaa Xaa Xaa Leu Xaa Xaa His Xaa Arg Thr His Thr Gly Glu Lys Pro
 50 55 60
 Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg
 65 70 75 80
 Lys Arg His Thr Lys Ile His Leu Arg Gln Lys Asp
 85 90

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TATGACTTGG ATGGGAGACC GCCTGG

26

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATTCCAGGC GGTCTCCCAT CCAAGTCA

28

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TATATAGCGT GGGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGTATATAC GCCCAGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TATATAGCGN NNGCGTATAT A

21

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGTATATAC GCNNCGCTA TATA

24

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTCCAATGGAG ACGCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTCCATGGAG ACGCAGGTGA GTTCCTCACG CCA

33

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCCTTTCTC TTCCAGAAGC CCTTCAGCGG CCA

33

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met	Ala	Glu	Glu	Lys	Pro	Phe	Gln	Cys	Arg	Ile	Cys	Met	Arg	Asn	Phe
1				5					10					15	

Ser	Asp	Arg	Ser	Ser	Leu	Thr	Arg	His	Thr	Arg	His	Thr	Gly	Glu	Lys
			20					25					30		

Pro

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met	Ala	Glu	Glu	Lys	Pro	Phe	Gln	Cys	Arg	Ile	Cys	Met	Arg	Asn	Phe
1				5					10					15	

Ser Glu Arg Gly Thr Leu Ala Arg His Glu Lys His Thr Gly Glu Lys
20 25 30

Pro

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Gly Gly Asn Leu
1 5 10 15

Val Arg His Leu Arg His Thr Gly Glu Lys Pro
20 25

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Gln Thr Leu
1 5 10 15

Gln Arg His Leu Lys His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Ala Thr Leu
 1 5 10 15
 Gln Arg His Leu Lys His Thr Gly Glu Lys
 20 25

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ala Gln Asp Leu
 1 5 10 15
 Gln Arg His Leu Lys His Thr Gly Glu Lys
 20 25

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Ala Glu Glu Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
 1 5 10 15
 Ser Asp Arg Ser Ser Leu Thr Arg His Thr Arg Thr His Thr Gly Glu
 20 25 30
 Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser
 35 40 45
 His Leu Thr Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln
 50 55 60
 Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Asn Leu Thr Arg
 65 70 75 80
 His Thr Arg Thr His Thr Gly Glu Lys
 85

Claims

1. A library of DNA sequences, each sequence encoding at least one zinc finger binding motif for display on a viral particle, the sequences coding for zinc finger binding motifs having random allocation of amino acids at positions -1, +2, +3, +6 and at least at one of positions +1, +5 and +8.
2. A library of DNA sequences, each sequence encoding the zinc finger binding motif of at least a middle finger of a zinc finger binding polypeptide for display on a viral particle, the sequence coding for the binding motif having random allocation of amino acids at positions -1, +2, +3 and +6.
3. A library of sequences according to claim 2, wherein the sequences coding for the binding motif have further random allocation of amino acids at one or more of positions +1, +5 and +8.
4. A library of sequences according to any one of claims 1, 2 or 3, wherein the sequences coding for the binding motif have random allocation of amino acids at positions +1, +5 and +8.
5. A library of sequences according to any one of the preceding claims, wherein the sequence encoded comprises a zinc finger polypeptide comprising a plurality of zinc fingers, adjacent fingers being joined by an intervening linker peptide.
6. A library of sequences according to any one of the preceding claims, wherein the sequence encoded comprises a zinc finger of the Zif 268 polypeptide.
7. A library of sequences according to any one of the preceding claims, wherein the sequence encoded comprises a zinc finger having random allocation of amino acids, positioned between two or more zinc fingers having a defined amino acid sequence.
8. A library of sequences according to any one of the preceding claims, in a form

suitable for cloning as a fusion with the minor coat protein of bacteriophage fd.

9. A method of designing a zinc finger polypeptide for binding to a particular target DNA sequence, comprising screening each of a plurality of zinc finger binding motifs against at least an effective portion of the target DNA sequence, and selecting those motifs which bind to the target DNA sequence.
10. A method according to claim 9, wherein two or more rounds of screening are performed.
11. A method of designing a zinc finger polypeptide for binding to a particular target DNA sequence, comprising comparing the binding of each of a plurality of zinc finger binding motifs to one or more DNA triplets, and selecting those motifs exhibiting preferable binding characteristics.
12. A method according to claim 11, further comprising an initial screening step according to claim 9 or 10.
13. A method of designing a zinc finger polypeptide for binding to a target DNA sequence, comprising combining in a single zinc finger polypeptide a plurality of zinc finger binding motifs, each of which has been screened by the method of claim 9 or 10, and/or selected by the method of claim 11 or 12.
14. A method according to claim 13, wherein the intervening linker peptide between adjacent zinc finger binding motifs is that present in a naturally occurring zinc finger binding polypeptide, or is an artificial peptide sequence, or is an artificial non-amino acid linker.
15. A zinc finger polypeptide for binding to a target DNA sequence, designed according to the method of any one of claims 9 to 14.
16. A DNA library consisting of 64 sequences, each sequence comprising a different one

of the 64 possible permutations of three DNA bases in a form suitable for use in the selection method of claim 11 or 12.

17. A library according to claim 16, wherein the sequences are associated, or are capable of being associated, with separation means.

18. A library according to claim 17, wherein the separation means is selected from one of the following: microtitre plate; magnetic or non-magnetic beads or particles capable of sedimentation; and an affinity chromatography column.

19. A library according to any one claims 16, 17 or 18, wherein the sequences are biotinylated.

20. A library according to any one of claims 16 to 19, wherein the sequences are contained within 12 mini-libraries.

21. A kit for making a zinc finger polypeptide for binding to a nucleic acid sequence of interest, comprising: a library of DNA sequences encoding zinc finger binding motifs of known binding characteristics in a form suitable for cloning into a vector; a vector molecule suitable for accepting one or more sequences from the library; and instructions for use.

22. A kit according to claim 21, wherein the vector is capable of directing the expression of the cloned sequences as a single zinc finger polypeptide.

23. A kit according to claim 21 or 22, wherein the vector is capable of directing the expression of the cloned sequences as a single zinc finger polypeptide displayed on the surface of a viral particle.

24. A kit for making a zinc finger polypeptide for binding to a nucleic acid sequence of interest, comprising: a library of DNA sequences, each encoding a zinc finger binding motif in a form suitable for screening according to the method of claim 9 or 10, and/or

- selecting according to the method of claim 11 or 12; and instructions for use.
25. A kit according to claim 24, wherein the library of DNA sequences is in accordance with any one of claims 1 to 8.
26. A kit according to claim 24 or 25, further comprising a library according to any one claims 16 to 20.
27. A kit according to any one claims 24, 25 or 26 further comprising appropriate buffer solutions and/or reagents for detection of bound zinc finger motifs.
28. A kit according to any one of claims 24 to 27, further comprising a vector suitable for accepting one or more sequences selected from the library of DNA sequences encoding zinc finger binding motifs.
29. A method of altering the expression of a gene of interest in a target cell, comprising: determining (if necessary) at least part of the DNA sequence of the structural region and/or a regulatory region of the gene of interest; designing a zinc finger polypeptide to bind to the DNA of determined sequence, and causing said zinc finger polypeptide to be present in the target cell.
30. A method according to claim 29, wherein the zinc finger polypeptide is designed in accordance with any one of claims 9-14.
31. A method according to claim 29 or 30, wherein the zinc finger polypeptide comprises one or more further functional domains.
32. A method according to any one of claims 29, 30 or 31, wherein the zinc finger polypeptide comprises a nuclear localisation signal so as to deliver the zinc finger polypeptide to the nucleus of the target cell.
33. A method according to any one of claims 29 to 32, wherein the zinc finger

polypeptide comprises the nuclear localisation signal from the large T antigen of SV40.

34. A method according to any one of claims 29 to 33, wherein the zinc finger polypeptide is caused to be present in the target cell by delivery into the cell of DNA directing the intracellular expression of the polypeptide.

35. A method of inhibiting cell division by altering the expression of a gene in accordance with the method of any one of claims 29 to 34, wherein the gene is one involved in regulating cell division.

36. A method of treating cancer, comprising delivering to a patient, or causing to be present therein, a zinc finger polypeptide which inhibits the expression of a gene enabling the cancer cells to divide.

37. A method of modifying a nucleic acid sequence of interest present in a sample mixture by binding thereto a zinc finger polypeptide, comprising contacting the sample mixture with a zinc finger polypeptide having affinity for at least a portion of the sequence of interest, so as to allow the zinc finger polypeptide to bind specifically to the sequence of interest.

38. A method according to claim 37, wherein the zinc finger polypeptide is designed in accordance with the method of any one of claims 9 to 14.

39. A method according to claim 37 or 38, further comprising the step of separating the zinc finger polypeptide (and nucleic acid sequences specifically bound thereto) from the rest of the sample.

40. A method according to any one of claims 37, 38 or 39, wherein the zinc finger polypeptide is bound to a solid phase support.

41. A method according to any one of claims 37 to 40, wherein the presence of the zinc finger polypeptide bound to the sequence of interest is detected by the addition of one or

more detection reagents.

42. A method according to any one of claims 37 to 41, wherein the DNA sequence of interest is present in an acrylamide or agarose gel matrix, or is present on the surface of a membrane.
43. A zinc finger polypeptide capable of inhibiting the expression of a disease-associated gene.
44. A zinc finger polypeptide according to claim 43, wherein the polypeptide is not naturally-occurring and is specifically designed to inhibit the expression of the disease-associated gene.
45. A zinc finger polypeptide according to claim 43 or 44, designed by the method of any one of claims 9 to 14.
46. A zinc finger polypeptide according to any one of claims 43, 44 or 45, capable of inhibiting the expression of an oncogene.
47. A zinc finger polypeptide according to any one of claims 43 to 46, capable of inhibiting the expression of a BCR-ABL fusion oncogene.
48. A zinc finger polypeptide according to any one of claims 43 to 47, designed to bind to the DNA sequence GCAGAAGCC.
49. A zinc finger polypeptide according to anyone of claims 43 to 46, capable of inhibiting the expression of a ras oncogene.
50. A zinc finger polypeptide according to claim 49, designed to bind to the DNA sequence GACGGCGCC.

1/17

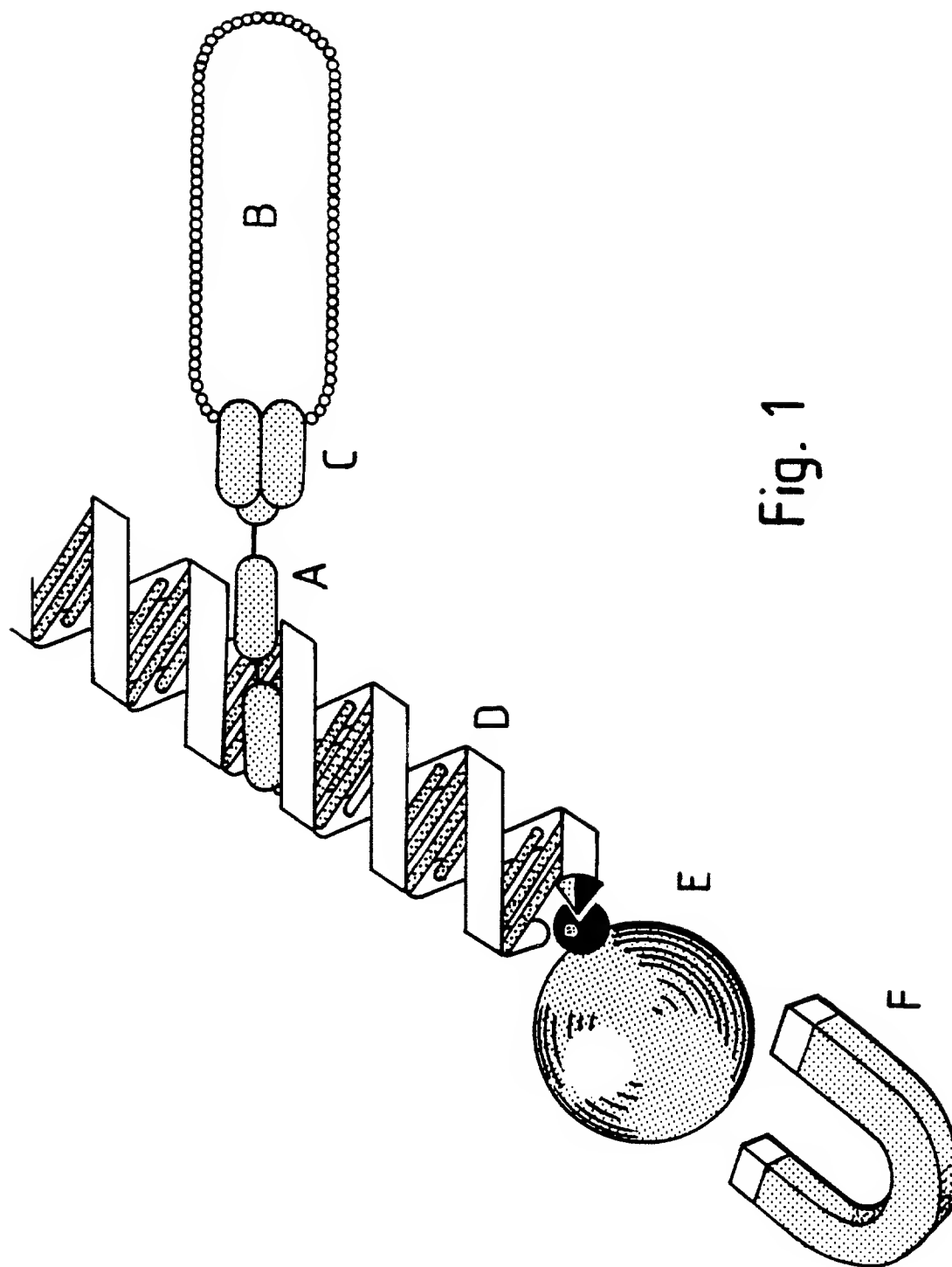
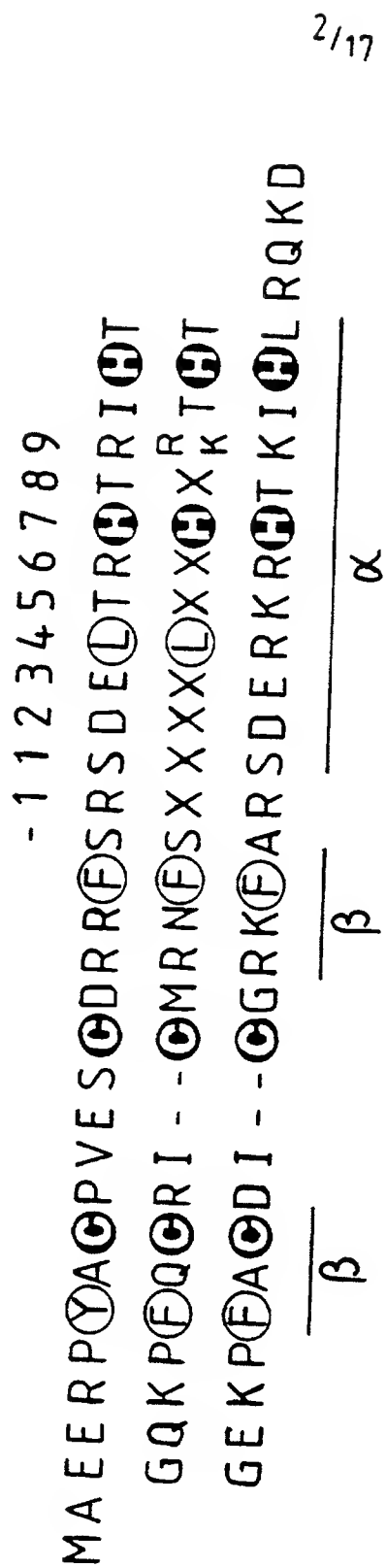


Fig. 1



2/17

Fig. 2

3/17

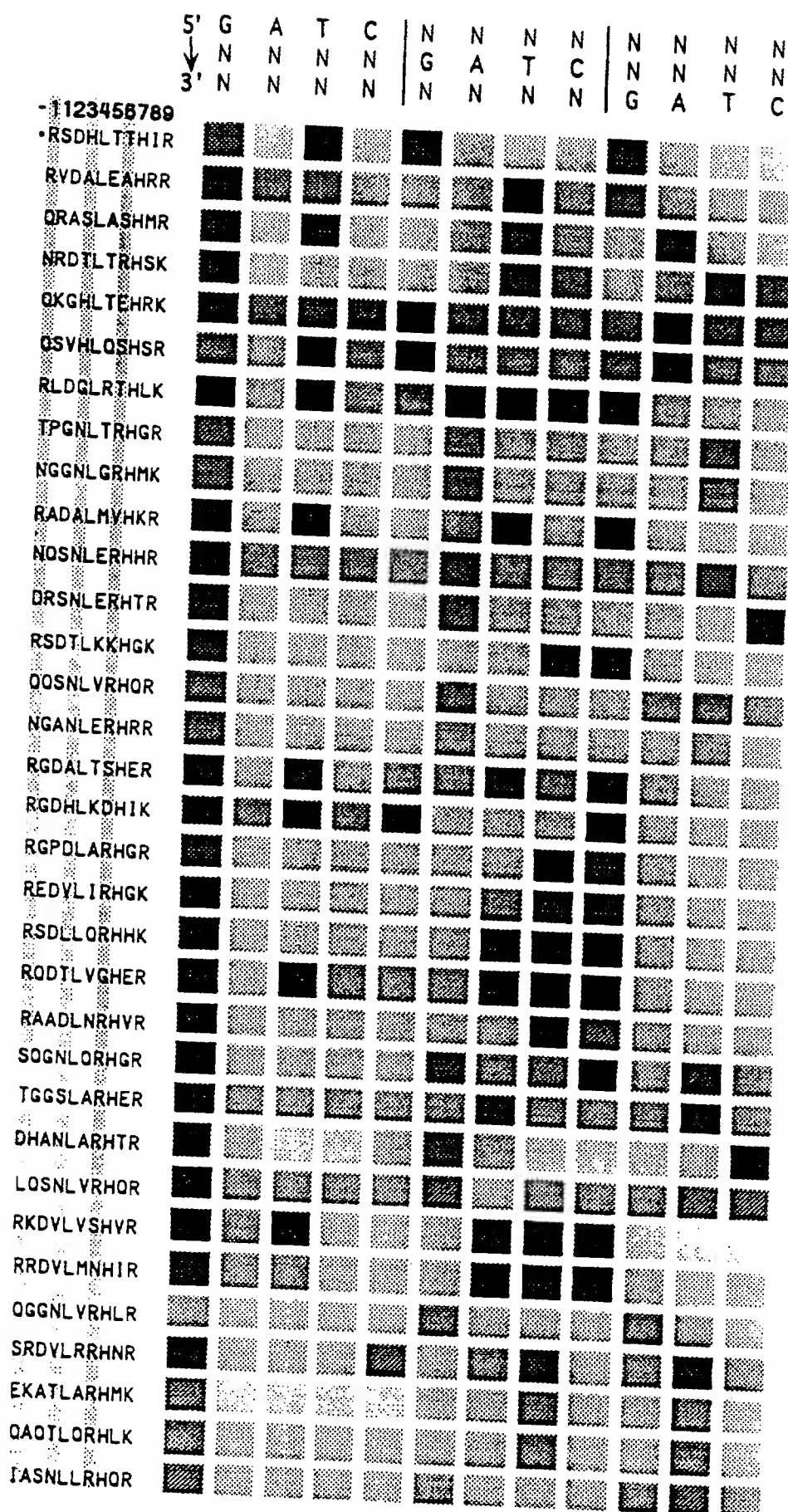
- (i) TATGACTTGGATGGAGACCGCCTGG
ACTGAACCTACCCCTCTGGCGGACCTTAA - (B)
- (ii) TATATACCCCTGGGGTATATA
ATATATCCACCCGATATATGCG - (B)
- (iii) TATATACCCGXXXXGGTATATA
ATATATCGGXXXXGGATATATGCG - (B)

Fig. 3

4/17

OLIGONUCLEOTIDE LIBRARY

Fig. 4

 α -HELIX SEQUENCE

5' → 3'

G/TGG

GTG

G/TTA

GTT

GGA

TGA

G/T^A/T/C^G

GAT

GAT

G/TTG

GAT

GAC

GCG

GAT

GAT

G/TTG

G/TGG

GCG

GT/C^GGT/C^GG/T^T/C^G

GCG

GAG/T

GTT

GAC

GAT/C

G/T^C/T^GGT/C^G

GAA

GCT

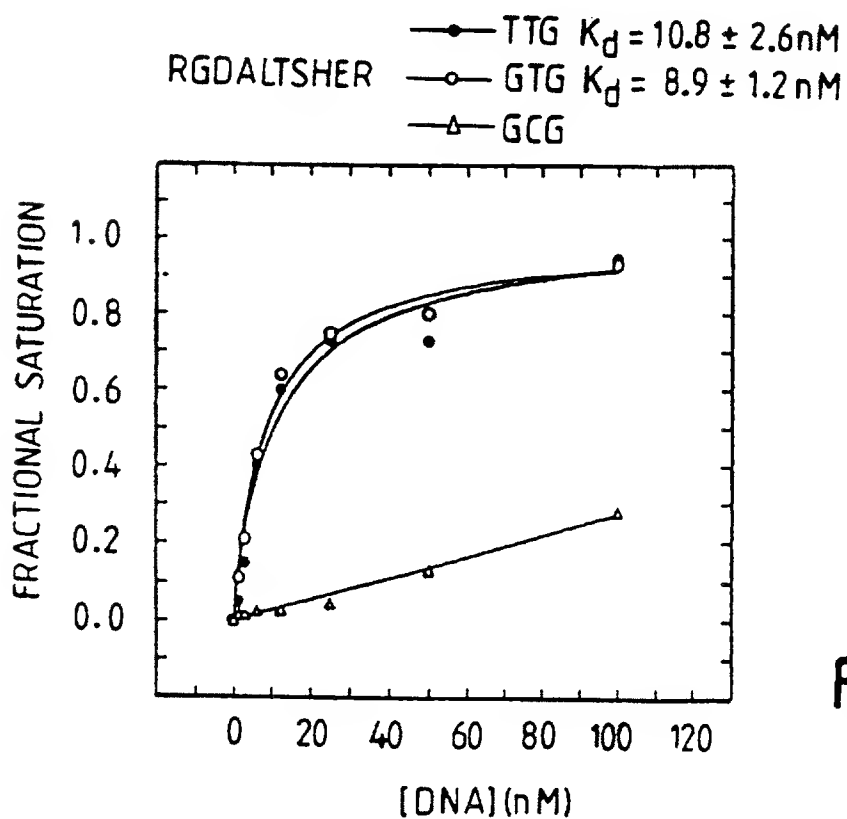
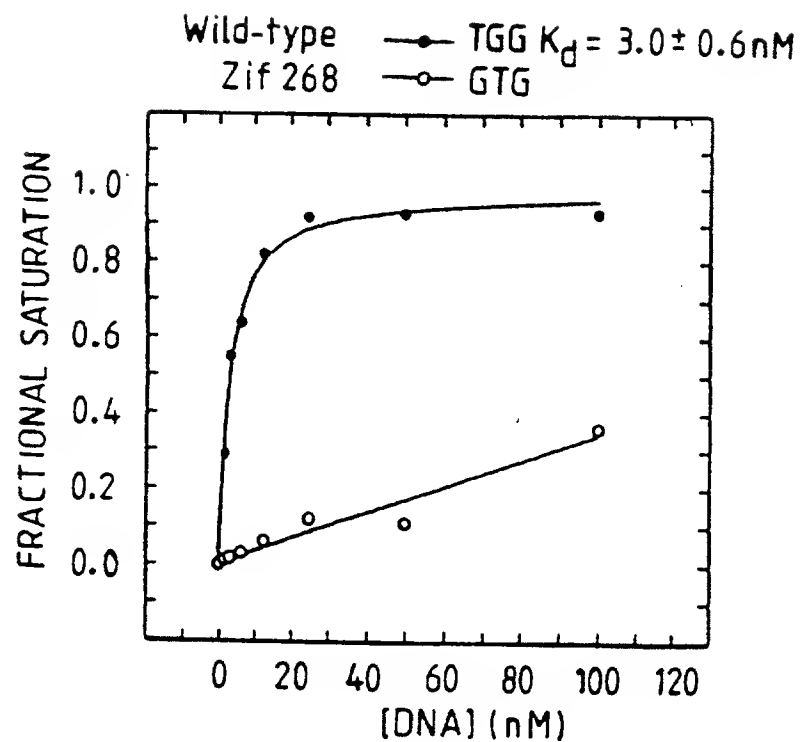
GCT

GCT

GAT/A/C

BINDING SITE SIGNATURE

5/17

Fig. 5
SHEET 1

6/17

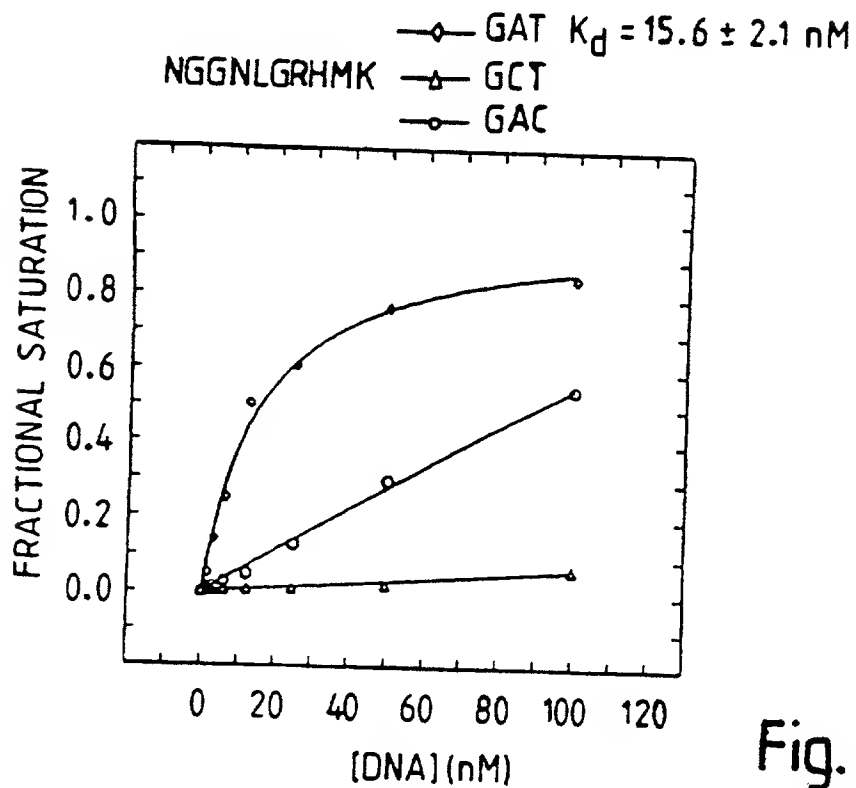
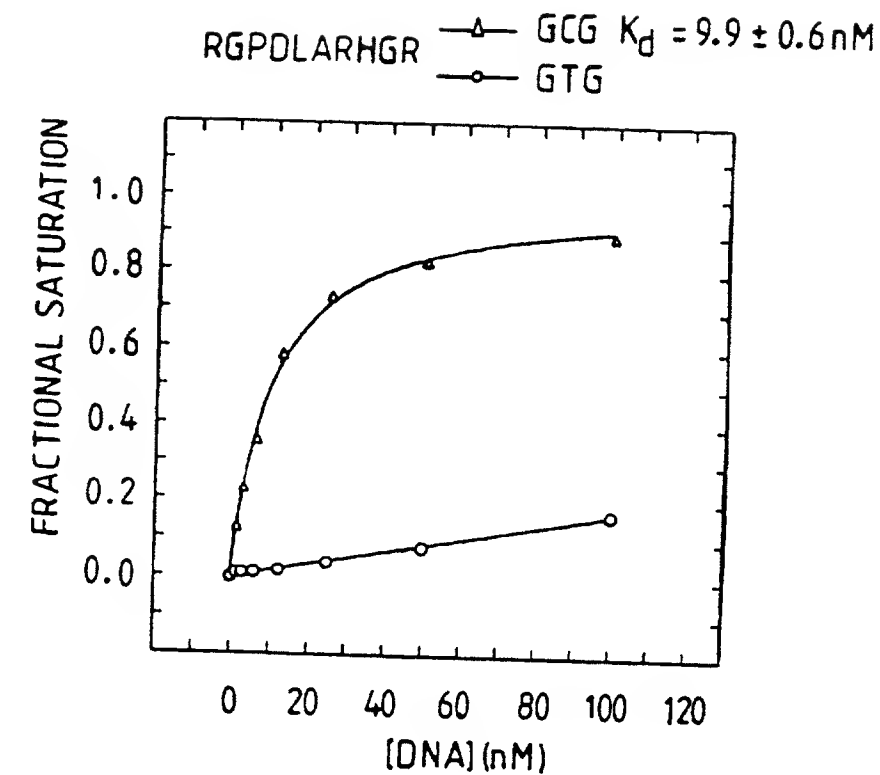
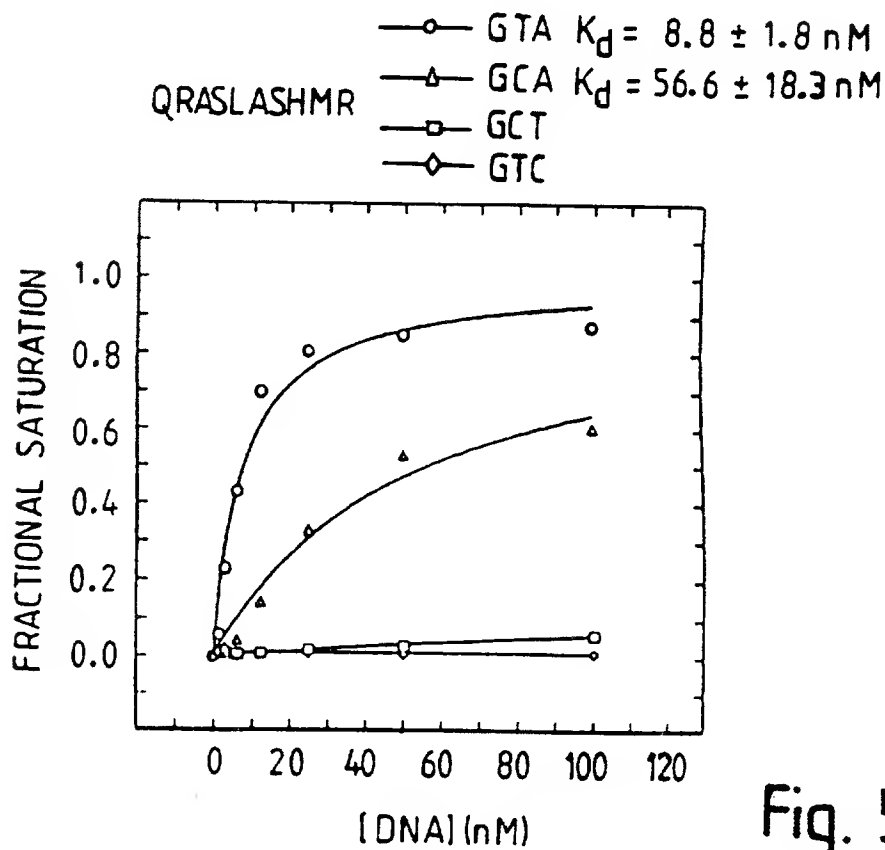
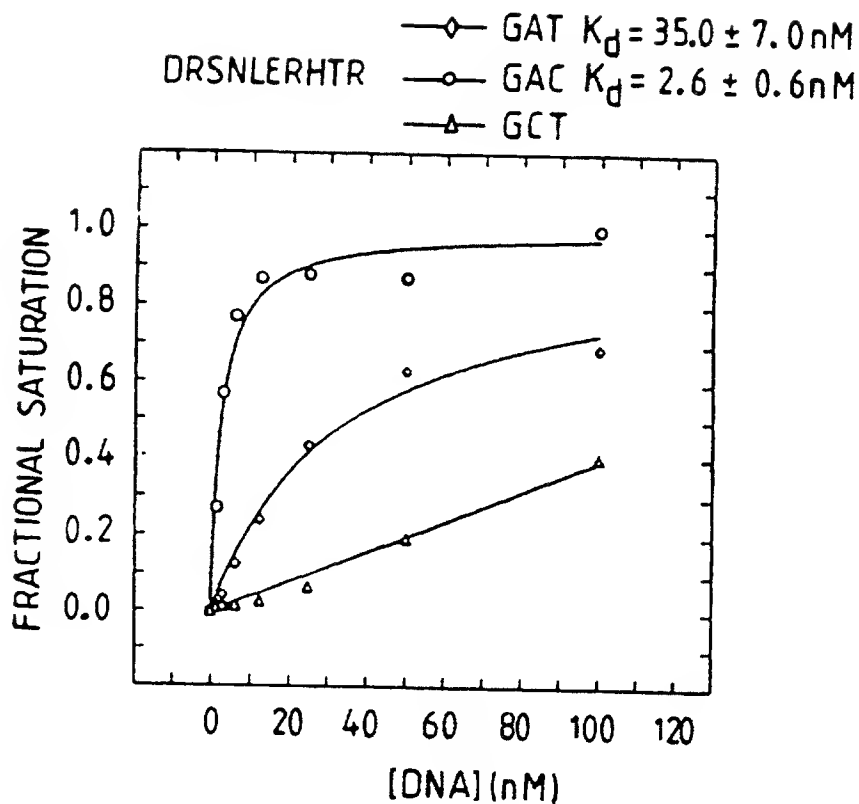


Fig. 5
SHEET 2

7/17

Fig. 5
SHEET 3

1 *BCR - ABL* TTC CAT GGA GAC GCA G AA GCC CTT CAG CGG CCA

2 *BCR* TTC CAT GGA GAC GCA G gt gag ttc ctc acg cca ^{8/17}

3 *ABL* ccc ctt tct ctt cca g AA GCC CTT CAG CGG CCA

Fig.6

9/17

		-1	1	2	3	4	5	6	7	8	9	
1A	M A E E K P F Q C R I C M R N F S D R S S L T R H T R H	T G E K P										
1B	M A E E K P F Q C R I C M R N F S E R G T L A R H E K H	T G E K P										
2A	F Q C R I C M R N F S Q G G N L V R H L R H	T G E K P										
3A	F Q C R I C M R N F S Q A Q T L Q R H L K H	T G E K										
3B	F Q C R I C M R N F S Q A A T L Q R H L K H	T G E K										
3C	F Q C R I C M R N F S Q A Q D L Q R H L K H	T G E K										
	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	
	β -SHEET	β -SHEET	α -HELIX						LINKER			

Fig. 7

10/17

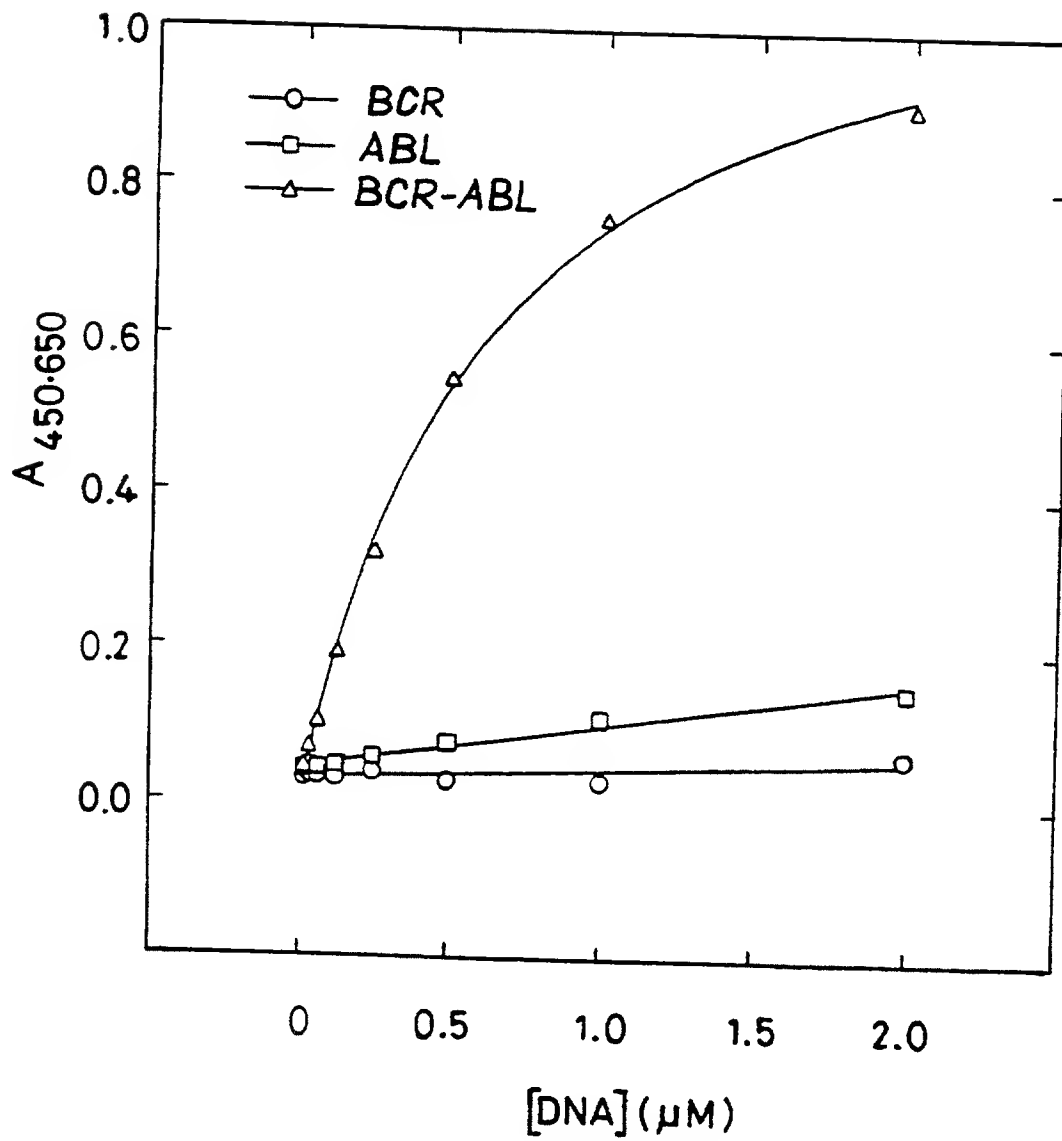


Fig. 8

11/17

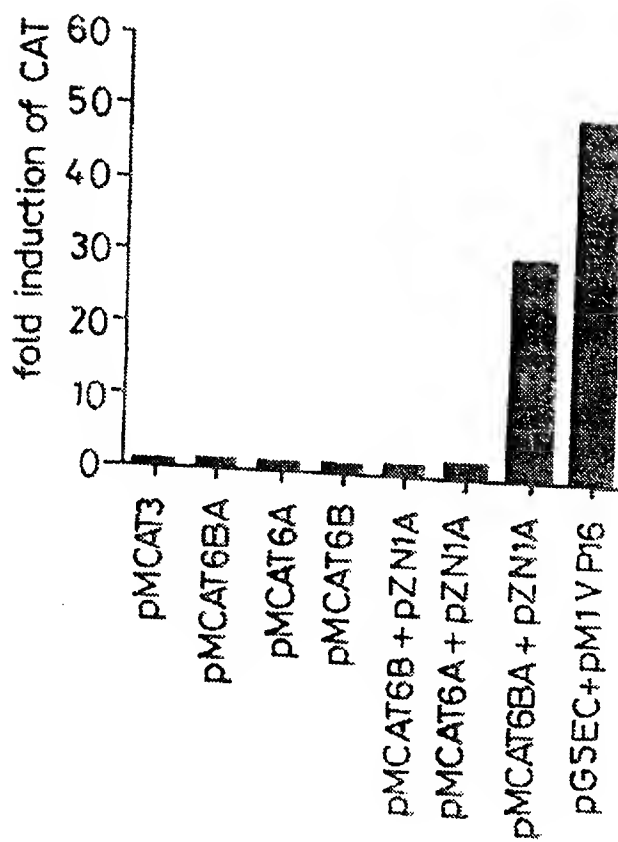
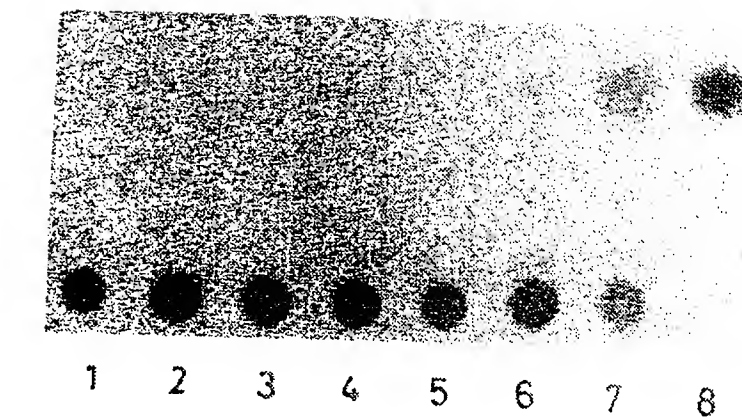
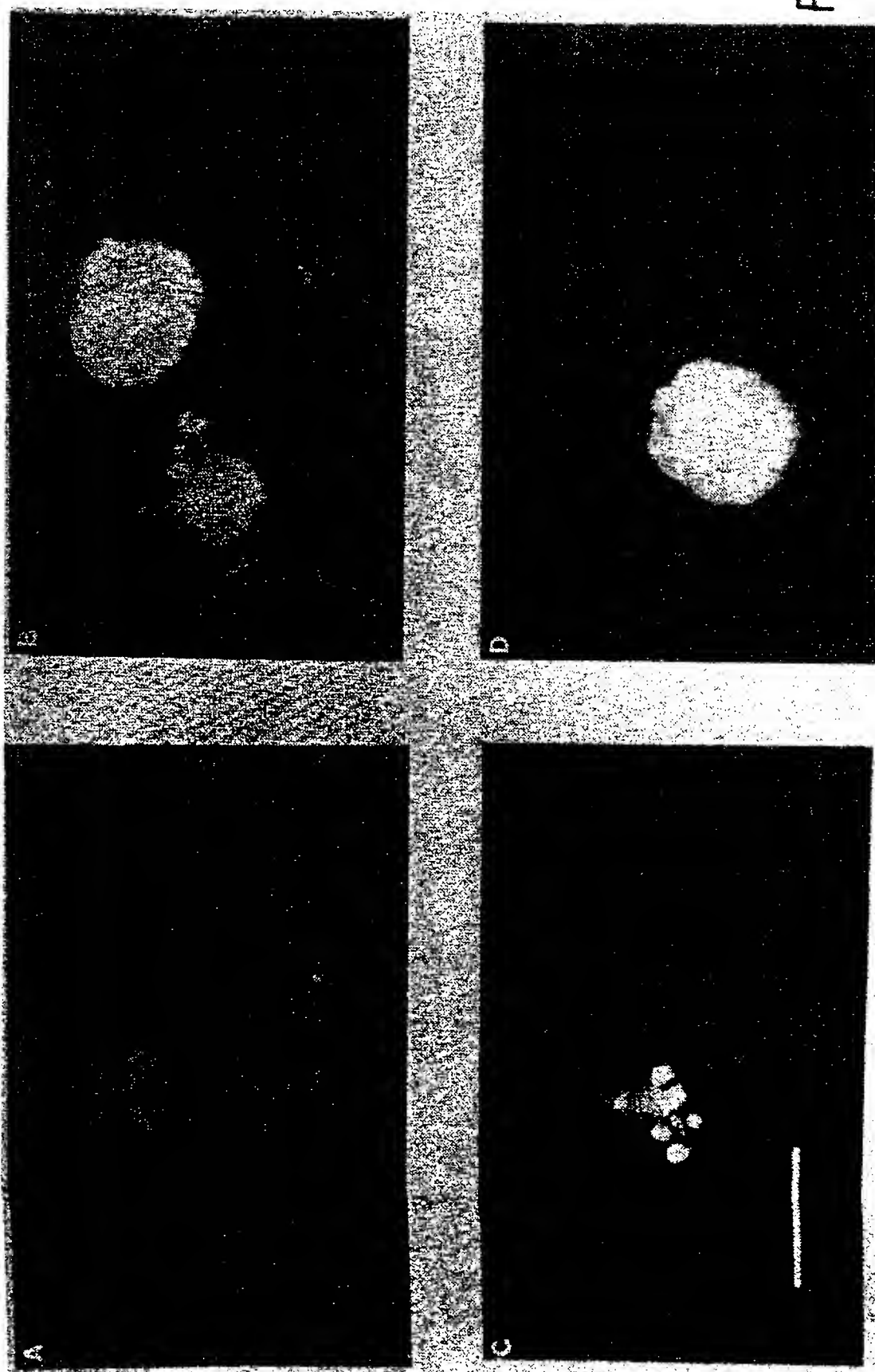


Fig. 9

12/17

Fig. 10



13/17

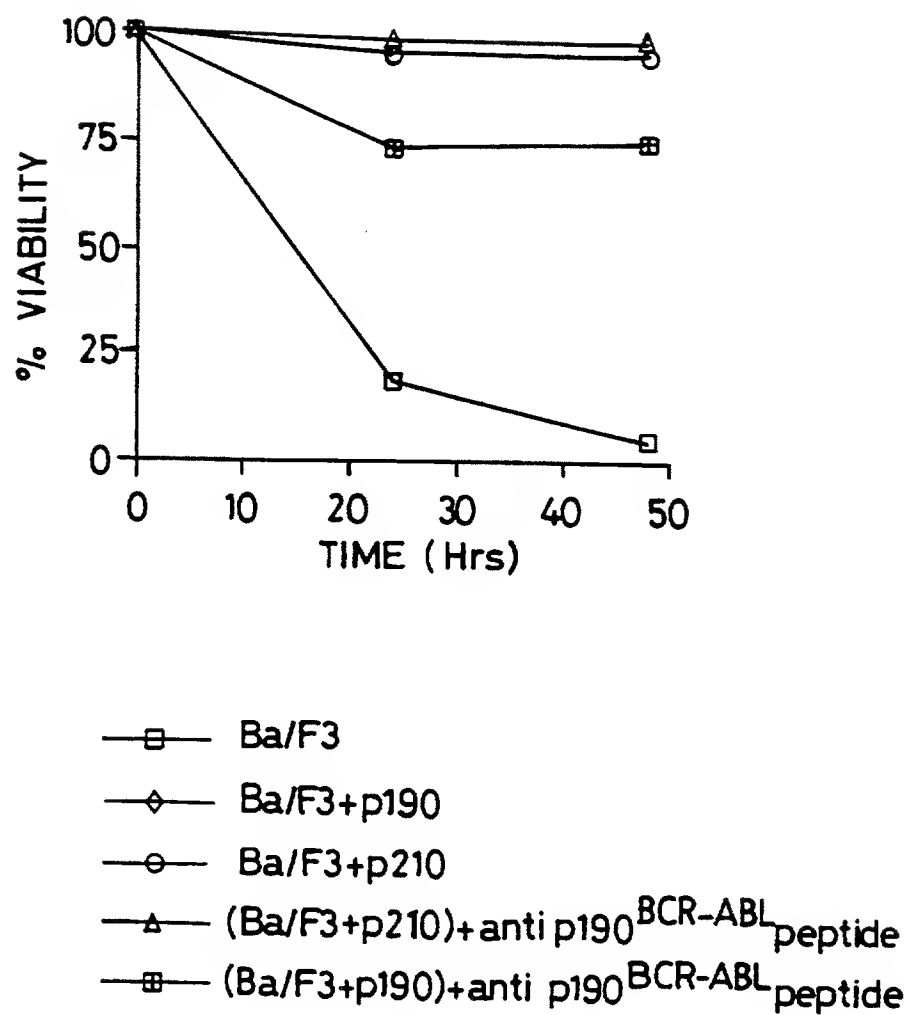


Fig. 11

14/17

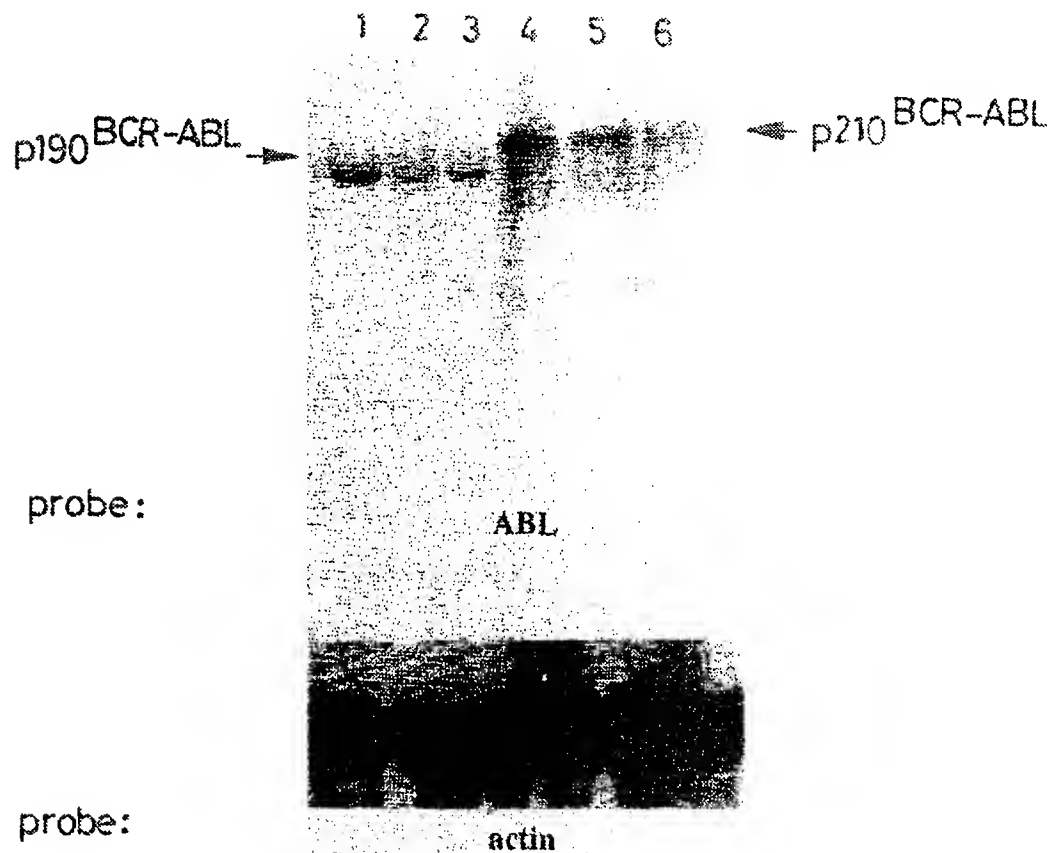


Fig. 12

15/17

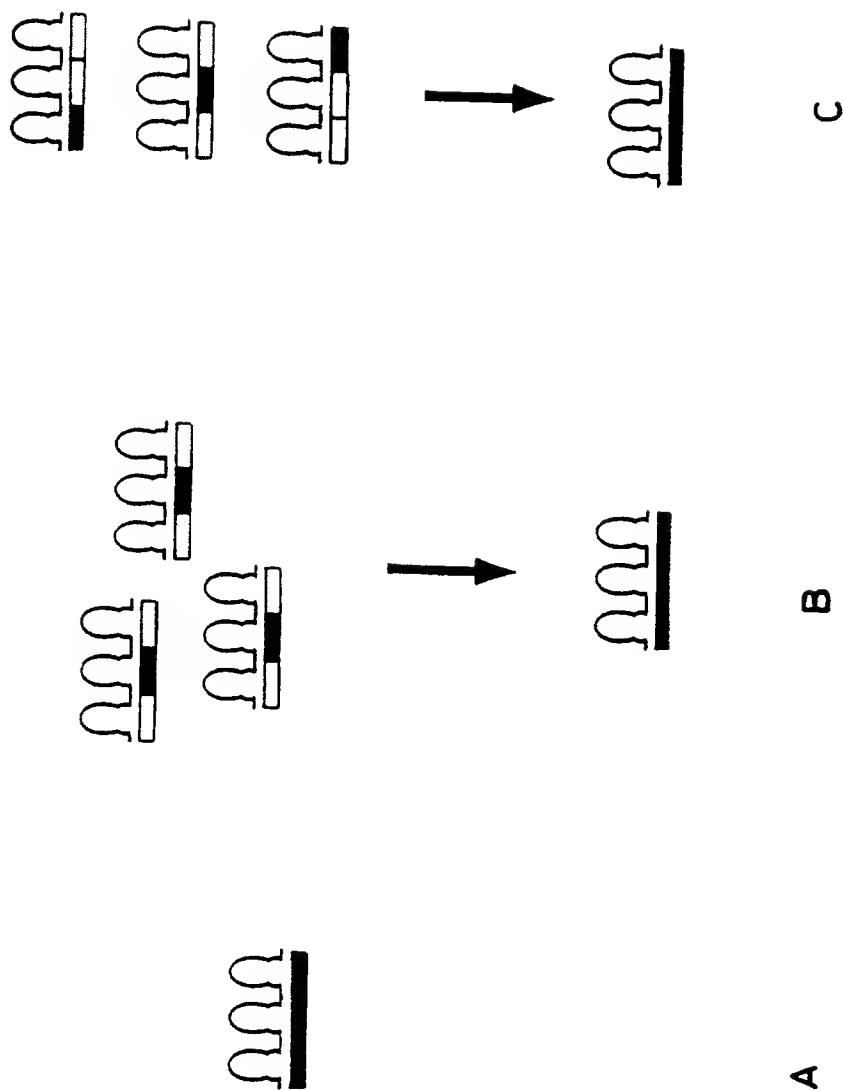


Fig. 13

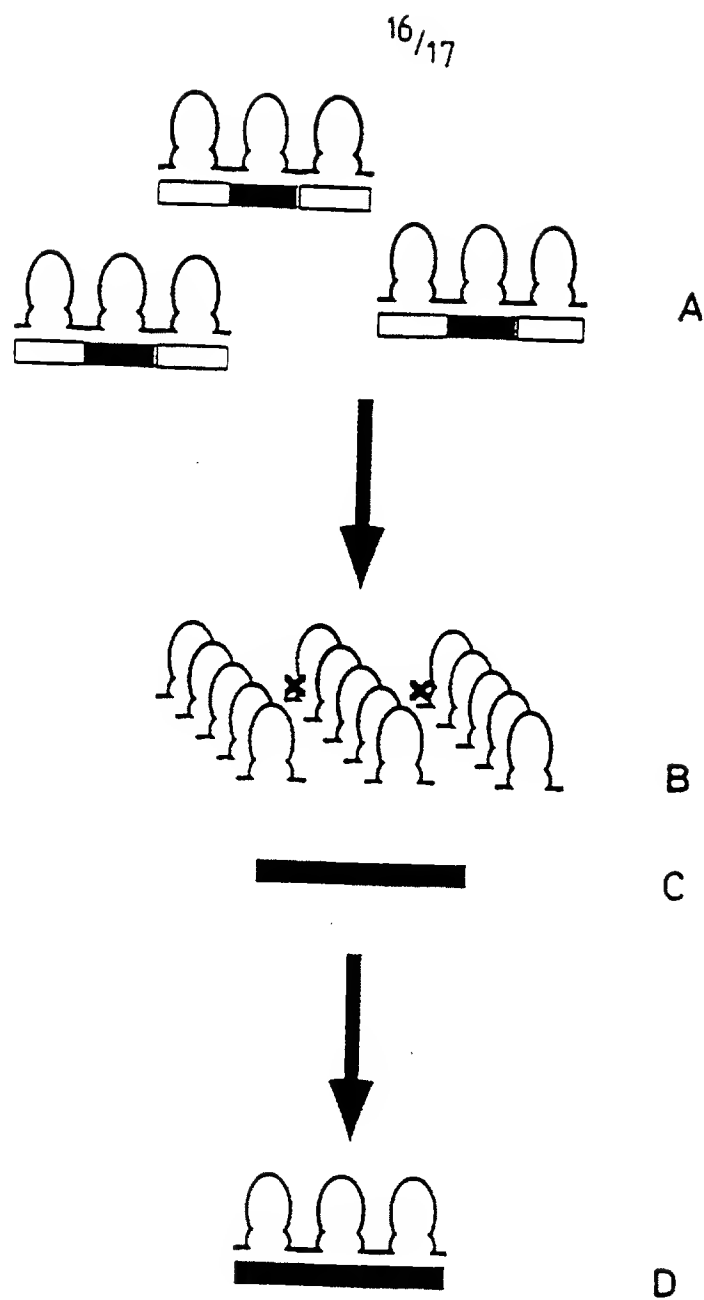


Fig. 14

17/17

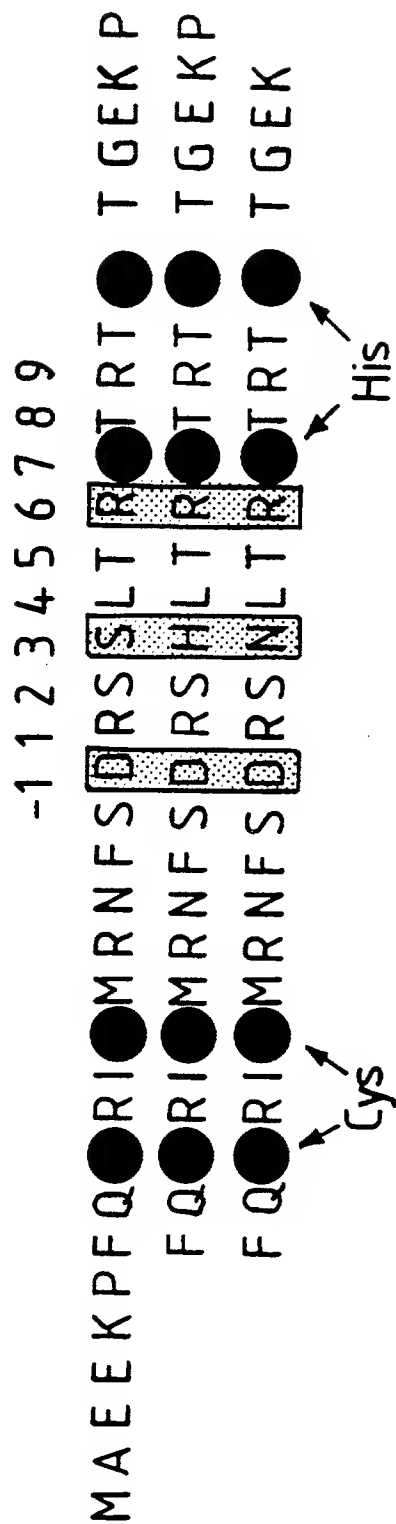


Fig. 15

INTERNATIONAL SEARCH REPORT

Int. Application No.

PCT/GB 95/01949

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/10 C12N15/12 C12N15/62 C12Q1/68 C07K14/47
A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12Q C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>SCIENCE, vol. 263, 4 February 1994 AAAS, WASHINGTON, DC, US, pages 671-673, E.J. REBAR AND C.O. PABO 'Zinc finger phage: Affinity selection of fingers with new DNA-binding specificities' cited in the application see the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1-30

☒ Further documents are listed in the continuation of box C.☐ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- *&* document member of the same patent family

Date of the actual completion of the international search

27 November 1995

Date of mailing of the international search report

08.12.95

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Hornig, H

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BIOCHEMISTRY, vol. 33, no. 19, 17 May 1994 AM. CHEM. SOC., WASHINGTON, DC, US, pages 5689-5695, A.C. JAMIESON ET AL. 'In vitro selection of zinc fingers with altered DNA-binding specificity' cited in the application see the whole document ---	1-30
A	FEBS LETTERS, vol. 283, no. 1, May 1991 ELSEVIER, AMSTERDAM, NL, pages 23-26, H.-J. THIESEN AND C. BACH 'Determination of DNA binding specificities of mutated zinc finger domains' cited in the application see the whole document ---	1-50
A	EMBO J., vol. 11, no. 12, December 1992 OXFORD UNIVERSITY PRESS, GB;, pages 4507-4517, G.H. JACOBS 'Determination of the base recognition positions of zinc fingers from sequence analysis' cited in the application see the whole document ---	1-50
A	PROC. NATL. ACAD. SCI., vol. 89, August 1992 NATL. ACAD SCI., WASHINGTON, DC, US;, pages 7345-7349, J.R. DESJARLAIS AND J.M. BERG 'Toward rules relating zinc finger protein sequences and DNA binding site preference' cited in the application see the whole document ---	1-50
A	NUCLEIC ACIDS RESEARCH, vol. 20, no. 16, 1992 IRL PRESS LIMITED, OXFORD, ENGLAND, pages 4137-4144, J. NARDELLI ET AL. 'Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis' cited in the application see the whole document ---	1-50

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>PROC. NATL.ACAD SCI., vol. 90, March 1993 NATL. ACAD SCI., WASHINGTON, DC, US;; pages 2256-2260, J.R. DESJARLAIS AND J.M. BERG 'Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins' cited in the application see the whole document</p> <p>---</p>	1-50
P,X	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 91 (23). 1994. 11163-11167. ISSN: 0027-8424, 8 November 1994 CHOO Y ET AL 'Toward a code for the interactions of zinc fingers with DNA: Selection of randomized fingers displayed on phage.' see the whole document</p> <p>---</p>	1-30
P,X	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 91 (23). 1994. 11168-11172. ISSN: 0027-8424, 8 November 1994 CHOO Y ET AL 'Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions.' see the whole document</p> <p>---</p>	1-30
P,X	<p>NATURE (LONDON) 372 (6507). 1994. 642-645. ISSN: 0028-0836, 15 December 1994 CHOO Y ET AL 'In vivo repression by a site-specific DNA- binding protein designed against an oncogenic sequence.' see the whole document</p> <p>---</p>	1-48
P,X	<p>PROC. NATL.ACAD SCI., vol. 92, no. 2, 17 January 1995 NATL. ACAD SCI., WASHINGTON, DC, US;; pages 344-348, H. WU ET AL. 'Building zinc fingers by selection: Toward a therapeutic application' cited in the application see the whole document</p> <p>---</p>	1-30

-/--

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 95/01949

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>FASEB JOURNAL 9 (8). 1995. 597-604. ISSN: 0892-6638, May 1995 KLUG A ET AL 'Zinc fingers.' see the whole document ---</p>	1-30
P,X	<p>CURRENT OPINION IN BIOTECHNOLOGY 6 (4). 1995. 431-436. ISSN: 0958-1669, August 1995 CHOO Y ET AL 'Designing DNA- binding proteins in the surface of filamentous phage.' see the whole document -----</p>	1-50

INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB95/01949

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claim 36 completely and (29-35 are partially as far as they concern an in vivo treatment) is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

1/17

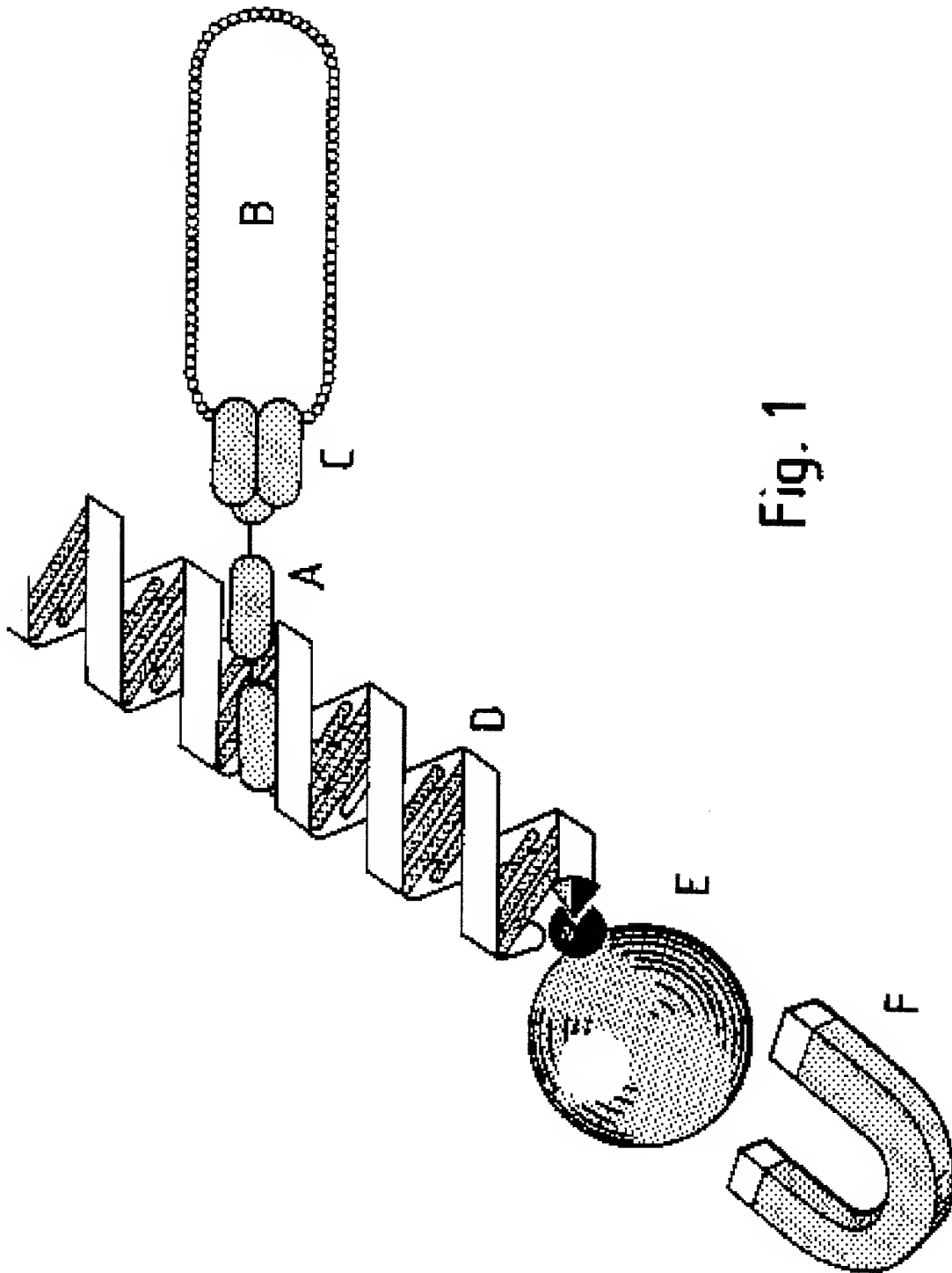
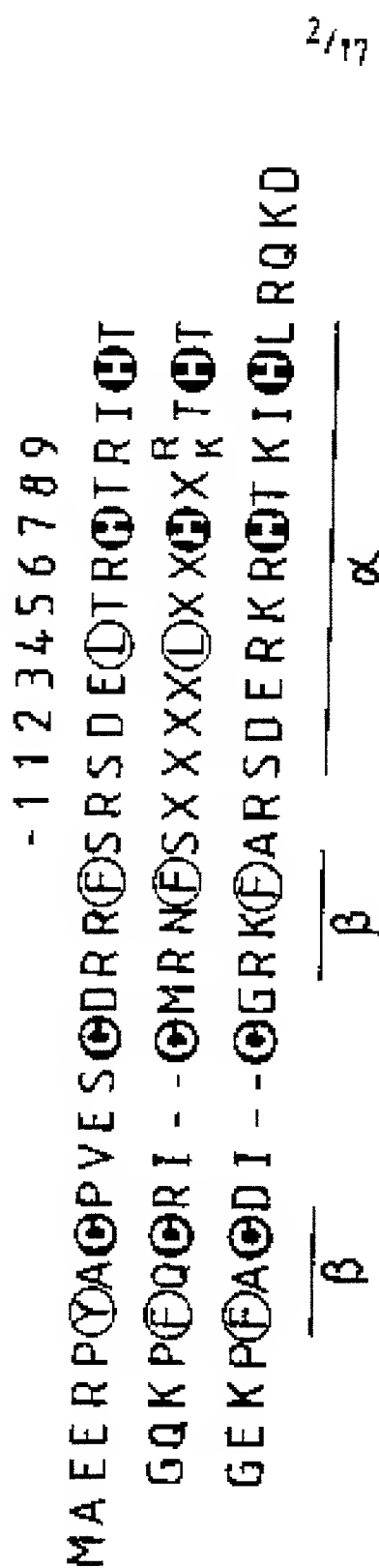


Fig. 1



2/17

Fig. 2

3/17

- (i) TATGACTTGGATGGGAGACCGCCTGG
ACTGAACCTACCCCTCTGGCGGACCTTAA - (B)
- (ii) TATATAGCCTGGGGTATATA
ATATATGGGACCGGATATATGCG - (B)
- (iii) TATATAGCCTGGGGTATATA
ATATATGGGACCGGATATATGCG - (B)

Fig. 3

4/17

OLIGONUCLEOTIDE LIBRARY

Fig. 4

α-HELIX SEQUENCE

	5' ↓ 3'	G N	A N	T N	C N	N G	N A	N T	N C	N G	N A	N T	N C
-1123456789													
RSOHLTHIR													
RYDALEAHR													
QRASLASHMR													
WRDILTHSRK													
QKGLTEHRK													
QSYHLSHR													
RLDGLRTHLK													
TPQNLTHGR													
NGGNLGRHMK													
RADALMHHR													
NOSNLERHHR													
QRSNLERHTR													
RSDILKKGK													
QOSNLVRHOR													
NGANLERKRR													
RSDALTSHER													
RGDHLKDHJK													
RGPDLARHGR													
REDVLIRHCK													
RSDDLGRHKK													
RDDTLVRHER													
RAADLNRHYR													
ROGNLOBHGR													
TGGLARHER													
DHANLARHTR													
LOSNLYRHOR													
RKDVLVSHVR													
RRDYLHKKIR													
QGGNLYBHLR													
SRDYLRRHVR													
EKATLARHMK													
QADTLORHLK													
LASNLLRHOR													

5' → 3'

5/7GG

GTG

5/7TA

GTT

GCA

TGA

5/7A/7/CG

GAT

GAT

5/7TG

GAT

GAC

GCG

GAT

GAT

5/7TG

5/7GG

GCG

GT/CG

GT/CG

5/7T/CG

GCG

GAC/7

GTT

GAC

GAT/CG

5/7C/7G

GT/CG

GAA

GCT

GCT

GCT

GAT/AC

BINDING SITE SIGNATURE

5/17

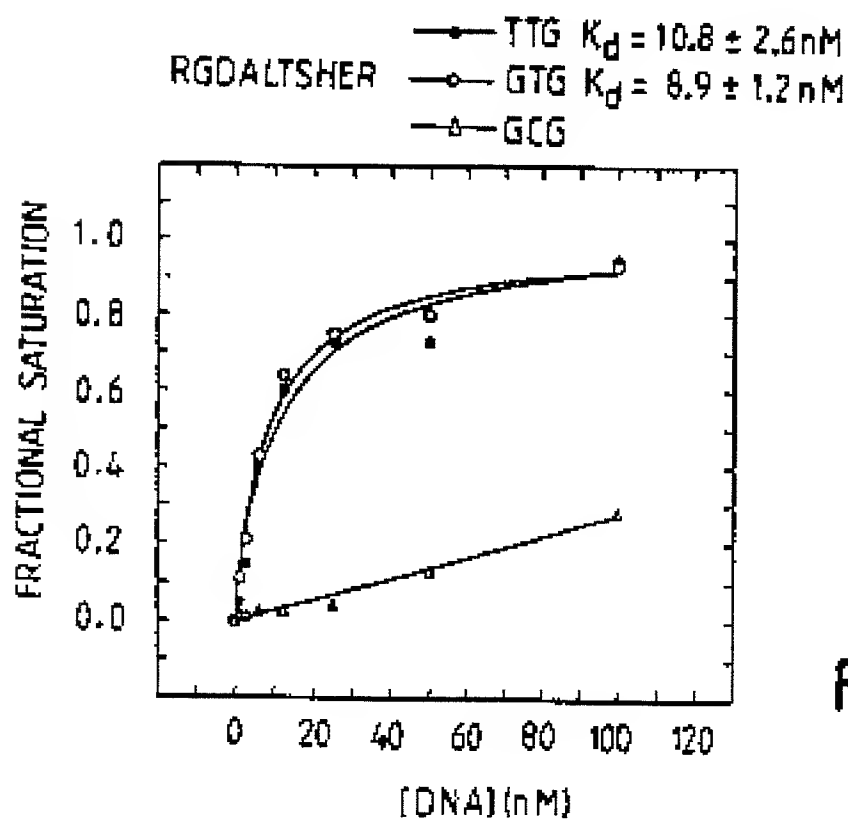
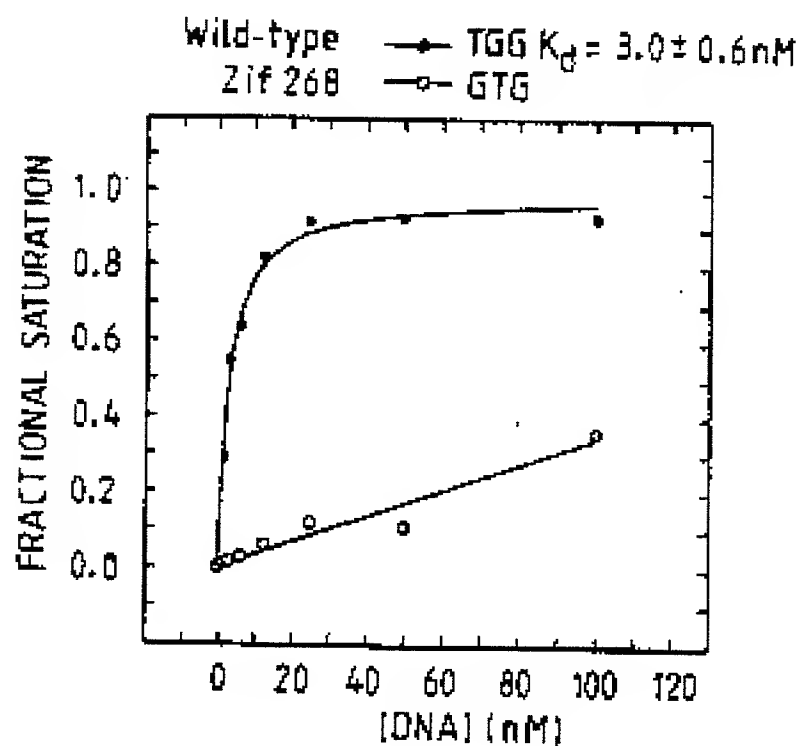


Fig. 5
SHEET 1

6/17

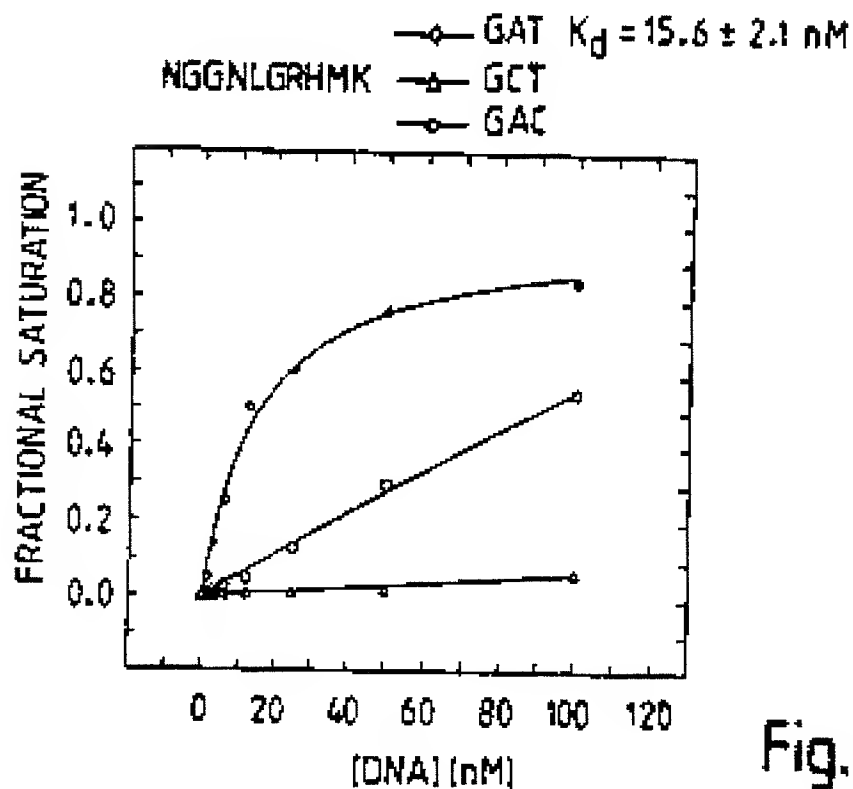
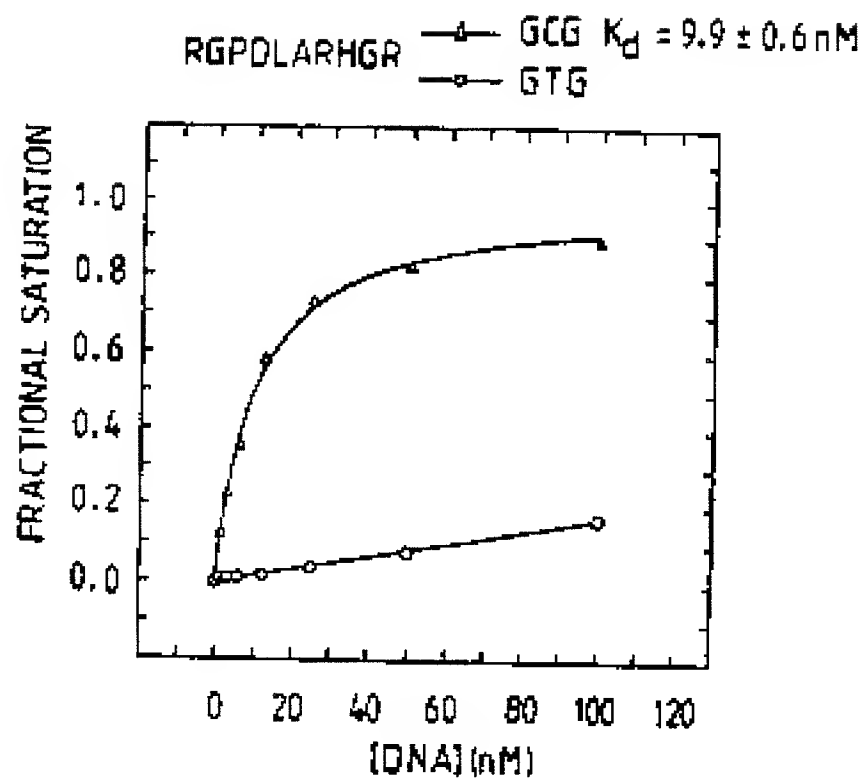


Fig. 5
SHEET 2

7/17

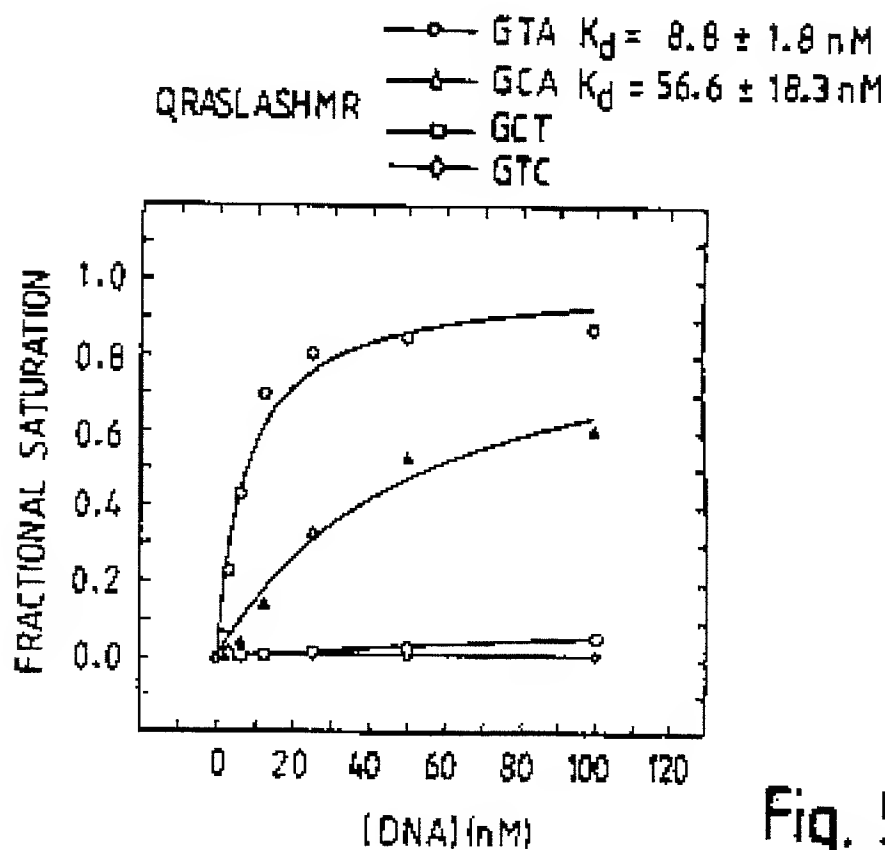
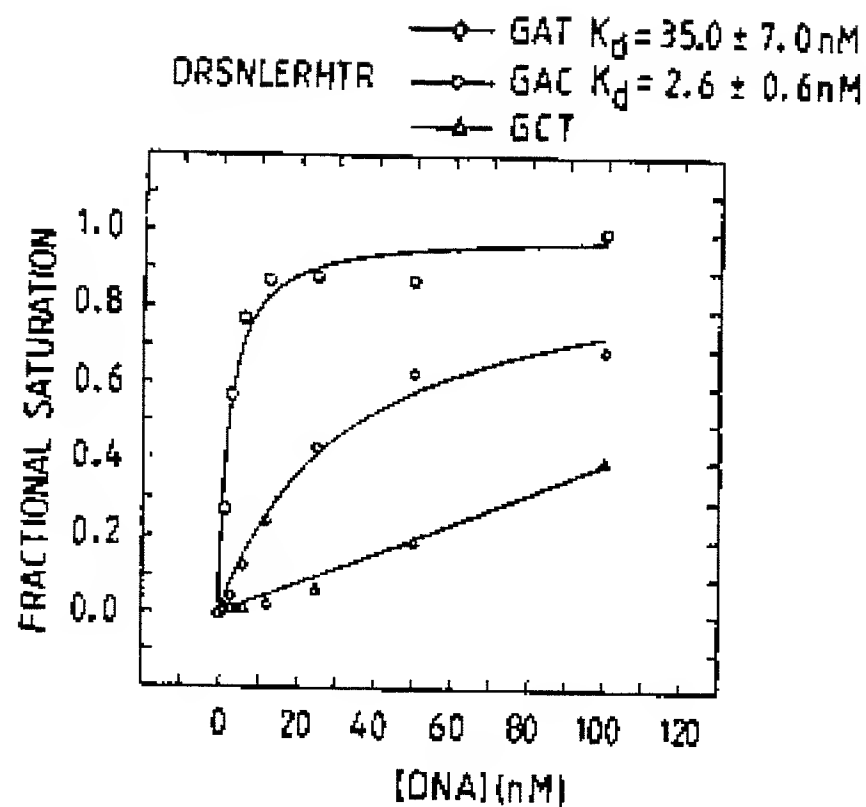


Fig. 5
SHEET 3

1 *BCR - ABL* TTC CAT GGA GAC GCA G AA GCC CTT CAG CGG CCA

2 *BCR* TTC CAT GGA GAC GCA G gt gag ttc ctc acg cca ^{8/17}

3 *ABL* ccc ctt tct ctt cca g AA GCC CTT CAG CGG CCA

Fig.6

	-1	1	2	3	4	5	6	7	8	9	
1A	M A E E K P F Q C R I C M R N F S D R S S L T R H T R H	T G E K P									
1B	M A E E K P F Q C R I C M R N F S E R G T L A R H E K H	T G E K P									
2A	F Q C R I C M R N F S Q G G N L V R H L R H	T G E K P									
3A	F Q C R I C M R N F S Q A Q T L Q R H L K H	T G E K									
3B	F Q C R I C M R N F S Q A A T L Q R H L K H	T G E K									
3C	F Q C R I C M R N F S Q A Q D L Q R H L K H	T G E K									
	<u>β-SHEET</u>	<u>β-SHEET</u>	<u>α-HELIX</u>							<u>LINKER</u>	

Fig. 7

10/17

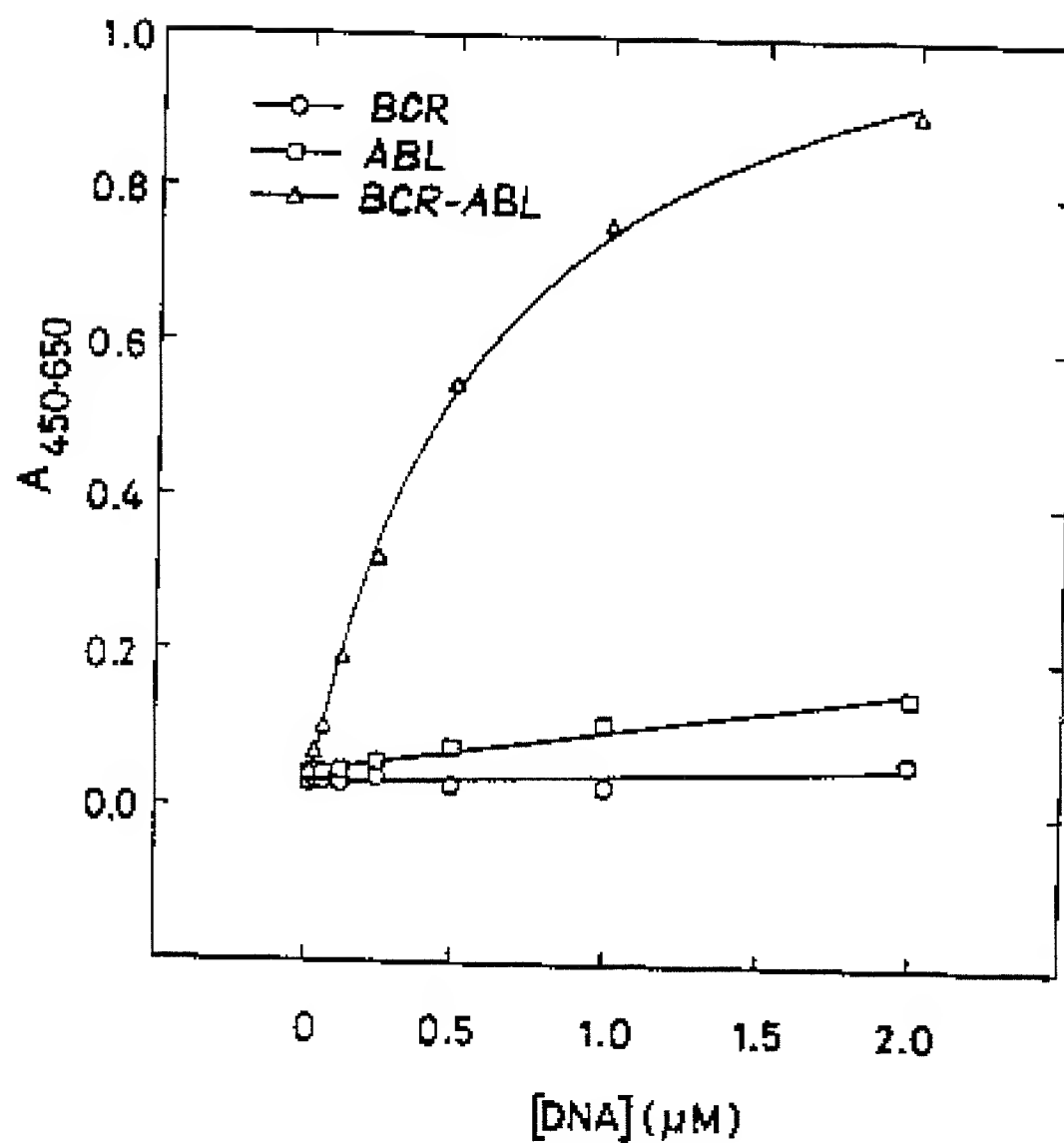


Fig. 8

11/17

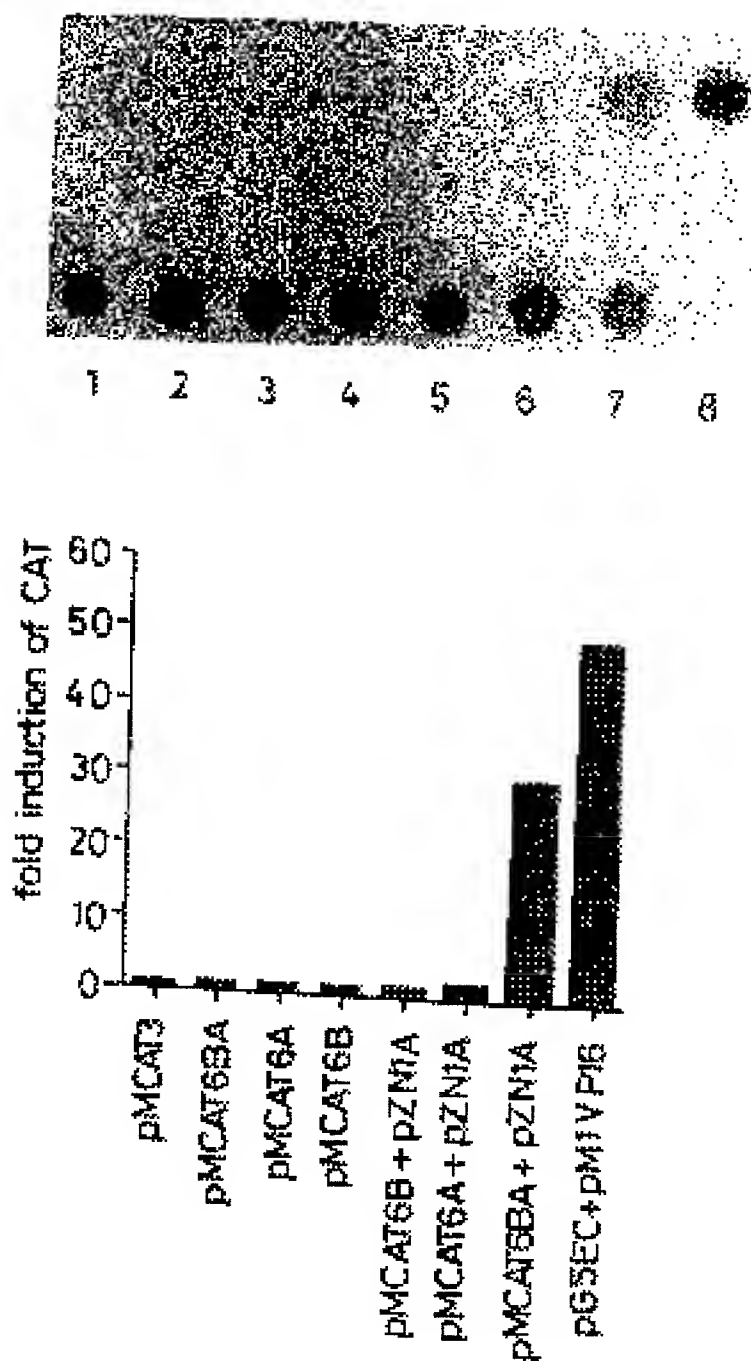
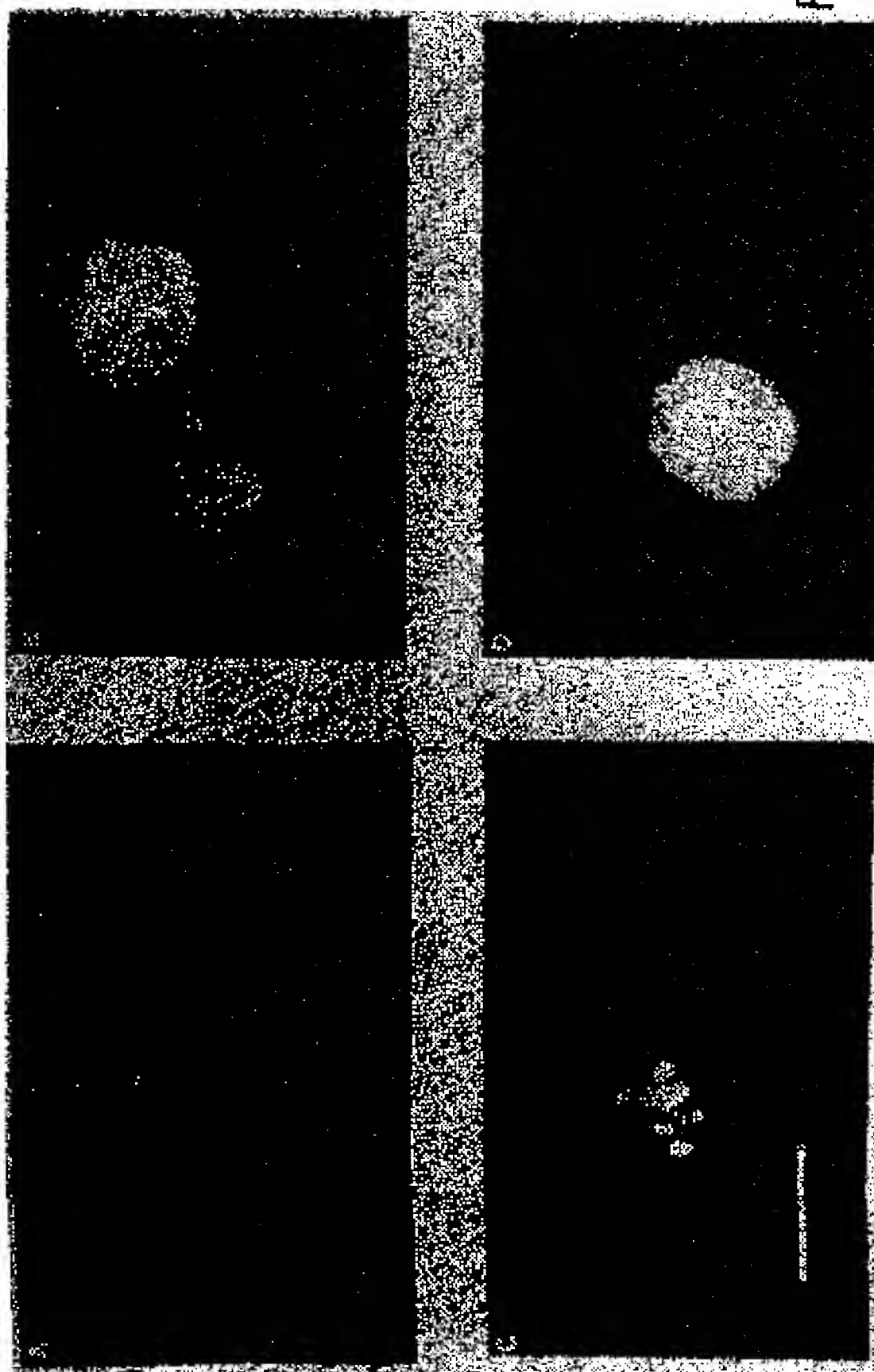


Fig. 9

12/17

Fig. 10



13/17

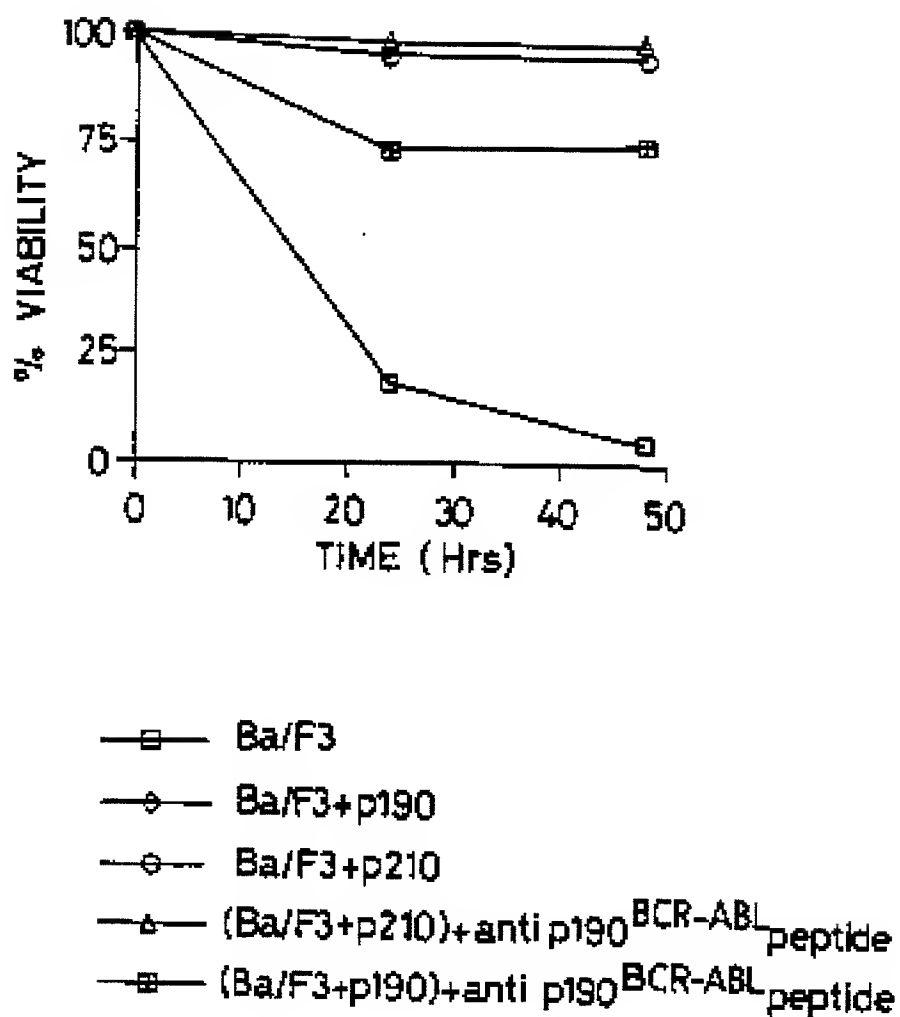


Fig. 11

16/17

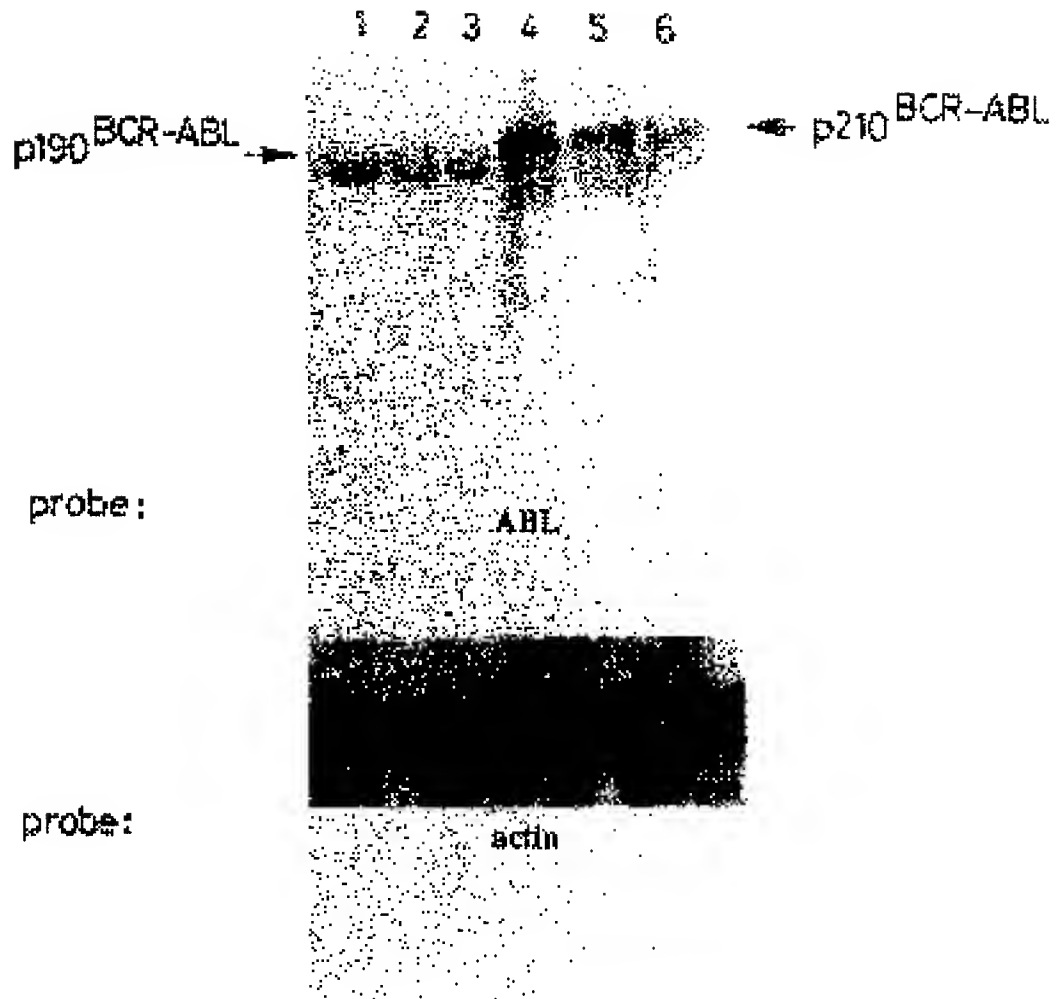


Fig. 12

15/17

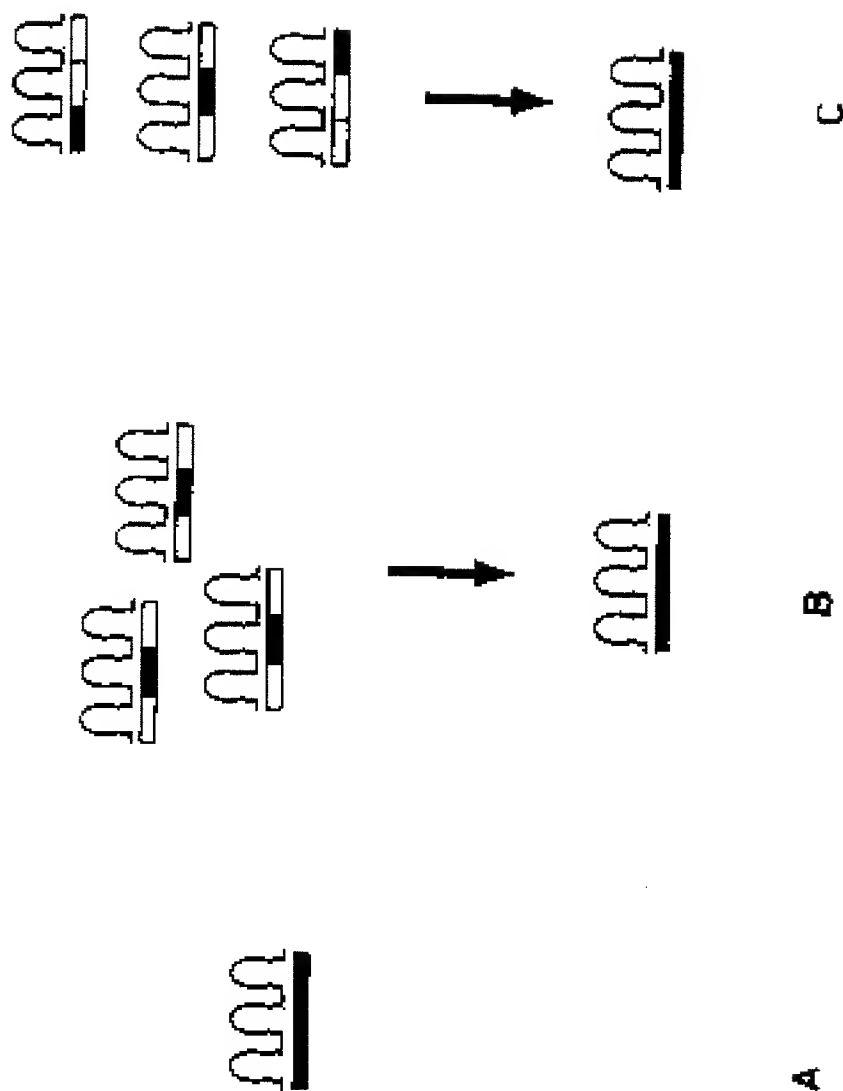


Fig. 13

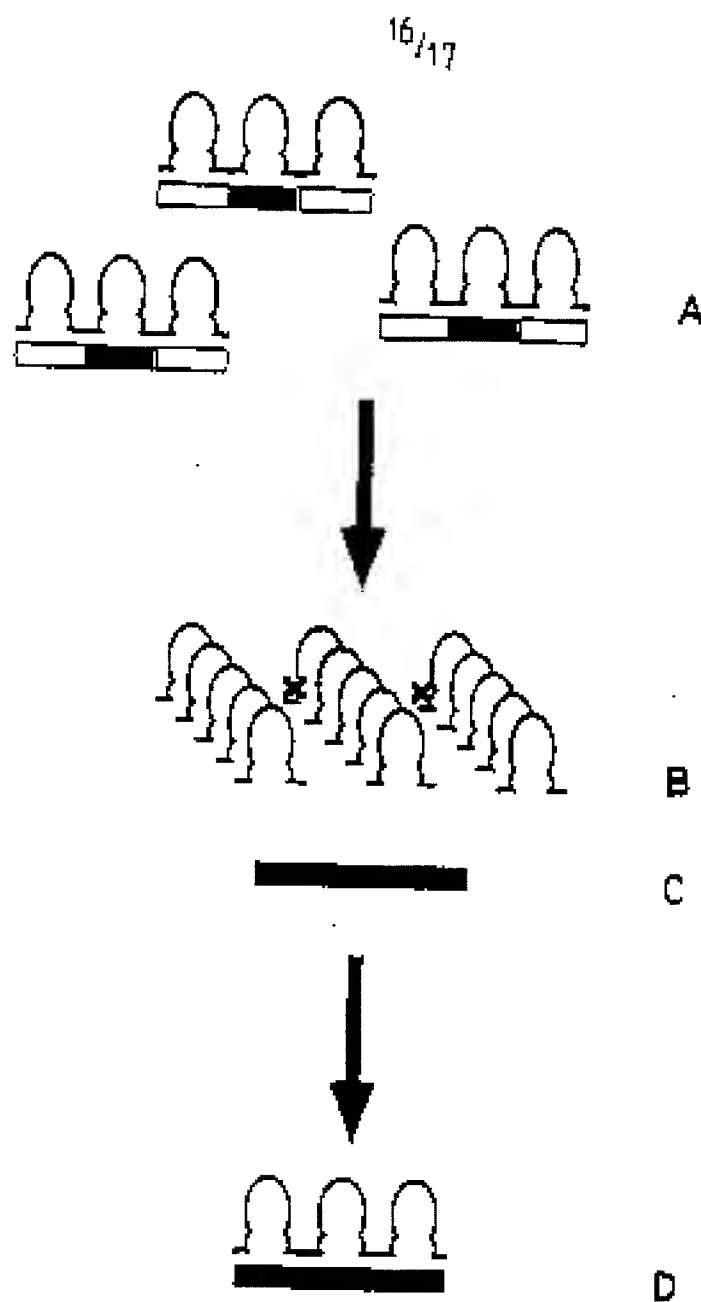


Fig. 14

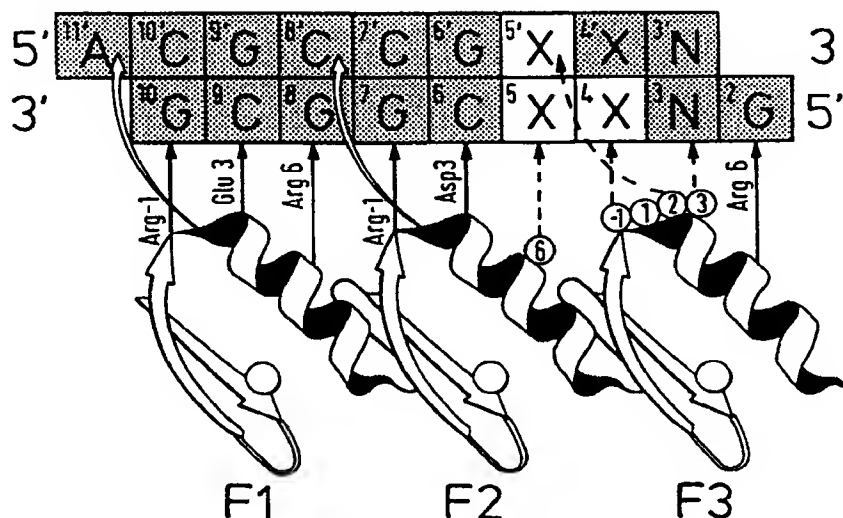


INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/10, 15/12, C07K 1/04, C12N 15/62, C12Q 1/68, C07K 14/47, A61K 48/00	A1	(11) International Publication Number: WO 98/53057 (43) International Publication Date: 26 November 1998 (26.11.98)
(21) International Application Number: PCT/GB98/01510 (22) International Filing Date: 26 May 1998 (26.05.98) (30) Priority Data: 9710809.6 23 May 1997 (23.05.97) GB (71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): CHOO, Yen [GR/GB]; MRC Laboratory of Molecular Biology, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH (GB). KLUG, Aaron [GB/GB]; MRC Laboratory of Molecular Biology, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH (GB). ISALAN, Mark [GB/GB]; 24 Shottfield Avenue, East Sheen, London SW14 8EA (GB). (74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: NUCLEIC ACID BINDING POLYPEPTIDE LIBRARY**(57) Abstract**

The invention relates to a zinc finger polypeptide library in which each polypeptide comprises more than one zinc finger which has been at least partially randomised, and to a set of zinc finger polypeptide libraries which encode overlapping zinc finger polypeptides, each polypeptide comprising more than one zinc finger which has been at least partially randomised, and which polypeptides may be assembled after selection to form a multifinger zinc finger polypeptide.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NUCLEIC ACID BINDING POLYPEPTIDE LIBRARY

The present invention relates to a library system for the selection of zinc finger polypeptides. In particular, the invention relates to a binary system, in which zinc finger motifs are randomised in overlapping regions and to smart libraries incorporating limited directed randomisation at selected positions.

Protein-nucleic acid recognition is a commonplace phenomenon which is central to a large number of biomolecular control mechanisms which regulate the functioning of eukaryotic and prokaryotic cells. For instance, protein-DNA interactions form the basis of the regulation of gene expression and are thus one of the subjects most widely studied by molecular biologists.

A wealth of biochemical and structural information explains the details of protein-DNA recognition in numerous instances, to the extent that general principles of recognition have emerged. Many DNA-binding proteins contain independently folded domains for the recognition of DNA, and these domains in turn belong to a large number of structural families, such as the leucine zipper, the "helix-turn-helix" and zinc finger families.

Despite the great variety of structural domains, the specificity of the interactions observed to date between protein and DNA most often derives from the complementarity of the surfaces of a protein α -helix and the major groove of DNA [Klug, (1993) Gene 135:83-92]. In light of the recurring physical interaction of α -helix and major groove, the tantalising possibility arises that the contacts between particular amino acids and DNA bases could be described by a simple set of rules; in effect a stereochemical recognition code which relates protein primary structure to binding-site sequence preference.

It is clear, however, that no code will be found which can describe DNA recognition by all DNA-binding proteins. The structures of numerous complexes show significant

differences in the way that the recognition α -helices of DNA-binding proteins from different structural families interact with the major groove of DNA, thus precluding similarities in patterns of recognition. The majority of known DNA-binding motifs are not particularly versatile, and any codes which might emerge would likely describe
5 binding to a very few related DNA sequences.

Even within each family of DNA-binding proteins, moreover, it has hitherto appeared that the deciphering of a code would be elusive. Due to the complexity of the protein-DNA interaction, there does not appear to be a simple "alphabetic" equivalence
10 between the primary structures of protein and nucleic acid which specifies a direct amino acid to base relationship.

International patent application WO 96/06166 addresses this issue and provides a "syllabic" code which explains protein-DNA interactions for zinc finger nucleic acid
15 binding proteins. A syllabic code is a code which relies on more than one feature of the binding protein to specify binding to a particular base, the features being combinable in the forms of "syllables", or complex instructions, to define each specific contact.

20 However, this code is incomplete, providing no specific instructions permitting the specific selection of nucleotides other than G in the 5' position of each triplet. The method relies on randomisation and subsequent selection in order to generate nucleic acid binding proteins for other specificities. Even with the aid of partial randomisation and selection, however, neither the method reported in WO 96/06166 nor any other
25 methods of the prior art have succeeded in isolating a zinc finger polypeptide based on the first finger of Zif268 capable of binding triplets wherein the 5' base is other than G or T. This is a serious shortfall in any ability to design zinc finger proteins.

Moreover, this document relies upon the notion that zinc fingers bind to a nucleic acid
30 triplet or multiples thereof, as does all of the prior art. We have now determined that zinc finger binding sites are determined by overlapping 4 bp subsites, and that

sequence-specificity at the boundary between subsites arises from synergy between adjacent fingers. This has important implications for the design and selection of zinc fingers with novel DNA binding specificities.

5 Summary of the Invention

The present invention recognises the importance of overlapping 4 bp subsite recognition in zinc finger polypeptide design. The resultant synergy between zinc fingers is overlooked in classical zinc finger library design, in which only a single zinc finger is
10 randomised in each library.

Accordingly, the present invention provides a zinc finger polypeptide library in which each polypeptide comprises more than one zinc finger which has been at least partially randomised.

15 Preferably, the invention provides a group of zinc finger polypeptide libraries which encode overlapping zinc finger polypeptides, each polypeptide comprising more than one zinc finger which has been at least partially randomised, and which polypeptides may be assembled after selection to form a multifinger zinc finger polypeptide.

20 In a further aspect, the invention relates to a library as described above in which randomisation is limited to substituting amino acids which are known to dictate variation in binding site specificity. The present invention provides a code of amino acid position bias which permits the selection of the library against any nucleic acid
25 sequence as the target sequence, and the production of a specific nucleic acid-binding protein which will bind thereto. Moreover, the invention provides a method by which a zinc finger protein specific for any given nucleic acid sequence may be designed and optimised. The present invention therefore concerns a recognition bias which has been elucidated for the interactions of classical zinc fingers with nucleic acid. In this case a
30 pattern of rules is provided which covers binding to all nucleic acid sequences.

The code set forth in the present invention takes account of synergistic interactions between adjacent zinc fingers, thereby allowing the selection of any desired binding site.

5

Brief Description of the Drawings

Figure 1 illustrates zinc finger-DNA interactions. A: model of classical triplet interactions with DNA base triplets in Zif268; B: similar model showing quadruplet interactions; C: model of library design for recognition code determination.

10

Figure 2 shows the amino acid sequence of three fingers used for phage display selection in the determination of recognition code.

Figure 3 lists the sequence-specific zinc finger clones obtained from phage selections, and their binding site signatures.

15

Figure 4 shows the base/amino acid correlation of the clones isolated from phage selections. Recognition patterns are highlighted.

20

Figure 5 illustrates the sequence-specific interactions selected for at position 2 of the α -helix, binding to position 1 of the quadruplet.

Figure 6 is a schematic diagram of the construction of a library according to the invention.

25

Detailed Description of the Invention

The present invention relates to libraries. The term "library" is used according to its common usage in the art, to denote a collection of polypeptides or, preferably, nucleic acids encoding polypeptides. The polypeptides of the invention contain regions of randomisation, such that each library will comprise or encode a repertoire of polypeptides, wherein individual polypeptides differ in sequence from each other. The same principle is present in virtually all libraries developed for selection, such as by phage display.

Randomisation, as used herein, refers to the variation of the sequence of the polypeptides which comprise the library, such that various amino acids may be present at any given position in different polypeptides. Randomisation may be complete, such that any amino acid may be present at a given position, or partial, such that only certain amino acids are present. Preferably, the randomisation is achieved by mutagenesis at the nucleic acid level, for example by synthesising novel genes encoding mutant proteins and expressing these to obtain a variety of different proteins. Alternatively, existing genes can be themselves mutated, such by site-directed or random mutagenesis, in order to obtain the desired mutant genes.

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the protein of interest. A number of methods for site-directed mutagenesis are known in the art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see "PCR Protocols: A guide to methods and applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York, 1990). Preferably, the commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the directions given by the manufacturer.

Screening of the proteins produced by mutant genes is preferably performed by expressing the genes and assaying the binding ability of the protein product. A simple

and advantageously rapid method by which this may be accomplished is by phage display, in which the mutant polypeptides are expressed as fusion proteins with the coat proteins of filamentous bacteriophage, such as the minor coat protein pII of bacteriophage m13 or gene III of bacteriophage Fd, and displayed on the capsid of bacteriophage transformed with the mutant genes. The target nucleic acid sequence is used as a probe to bind directly to the protein on the phage surface and select the phage possessing advantageous mutants, by affinity purification. The phage are then amplified by passage through a bacterial host, and subjected to further rounds of selection and amplification in order to enrich the mutant pool for the desired phage and eventually isolate the preferred clone(s). Detailed methodology for phage display is known in the art and set forth, for example, in US Patent 5,223,409; Choo and Klug, (1995) Current Opinions in Biotechnology 6:431-436; Smith, (1985) Science 228:1315-1317; and McCafferty *et al.*, (1990) Nature 348:552-554; all incorporated herein by reference. Vector systems and kits for phage display are available commercially, for example from Pharmacia.

The polypeptides which comprise the libraries according to the invention are zinc finger polypeptides. In other words, they comprise a Cys2-His2 zinc finger motif. It is a feature of the invention that each polypeptide comprises more than one zinc finger, such that the library may be selected on the basis of the interaction between two or more zinc fingers on the polypeptide.

Zinc fingers, as is known in the art, are nucleic acid binding molecules. Each zinc finger binds to a quadruplet sequence in a target nucleic acid through contacts between specific amino acid residues of the α -helix of the zinc finger and the nucleic acid strand. The quadruplets specified in the present invention are overlapping, such that, when read 3' to 5' on the -strand of the nucleic acid, base 4 of the first quadruplet is base 1 of the second, and so on. Accordingly, in the present application, the bases of each quadruplet are referred by number, from 1 to 4, 1 being the 3' base and 4 being the 5' base. Base 4 is equivalent to the 5' base of a classical zinc finger binding triplet. In general, base 4 is bound through a contact at position +6 of the α -helix, base 3

through a contact at position +3, base 2 through a contact at position -1 and base 1 through a contact to the opposite strand of double-stranded nucleic acids at position +2.

- 5 All of the nucleic acid-binding residue positions of zinc fingers, as referred to herein, are numbered from the first residue in the α -helix of the finger, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α -helix in a Cys2-His2 zinc finger polypeptide.
- 10 Residues referred to as "+ +2" are residues present in an adjacent (C-terminal) finger. They reflect the synergistic cooperation between position +2 on base 1 (on the + strand) and position +6 of the preceding (N-terminal) finger on base 4 of the preceding (3') quadruplet, which is the same base due to the overlap. Where there is no C-terminal adjacent finger, "+ +" interactions do not operate.
- 15 Cys2-His2 zinc finger binding proteins, as is well known in the art, bind to target nucleic acid sequences via α -helical zinc metal atom co-ordinated binding motifs known as zinc fingers. Each zinc finger in a zinc finger nucleic acid binding protein is responsible for determining binding to a nucleic acid quadruplet in a nucleic acid
- 20 binding sequence. Preferably, there are 2 or more zinc fingers, for example 2, 3, 4, 5 or 6 zinc fingers, in each binding protein. Advantageously, there are 3 zinc fingers in each zinc finger binding protein.

The present invention allows the production of what are essentially artificial nucleic

25 acid binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover,

30 any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein

therefore specifically comprises within its scope functional analogues of the defined amino acids.

The α -helix of a zinc finger binding protein aligns antiparallel to the nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N terminal to C-terminal sequence of the zinc finger. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus to C-terminus, the result is that when a nucleic acid sequence and a zinc finger protein are aligned according to convention, the primary interaction of the zinc finger is with the - strand of the nucleic acid, since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain fingers, such as finger 4 of the protein GLI, bind to the + strand of nucleic acid: see Suzuki *et al.*, (1994) NAR 22:3397-3405 and Pavletich and Pabo, (1993) Science 261:1701-1707. The incorporation of such fingers into nucleic acid binding molecules according to the invention is envisaged.

The libraries of the present invention allow selection for synergistic cooperation between adjacent zinc fingers by promoting coselection of adjacent fingers against a single DNA target. This is achieved by randomising, in the same zinc finger polypeptide, more than one zinc finger. In a preferred embodiment, approximately one and a half zinc fingers are randomised in each polypeptide, but this may be varied according to library design.

The zinc finger polypeptides encoded in the library of the invention may comprise any number of zinc fingers, provided this is more than one. Advantageously, each polypeptide encodes between three and six zinc fingers. In each library, the randomisation extends to cover the overlap of at least one pair of zinc fingers. Preferably, the overlap of a single pair is covered.

Preferably, the libraries of the present invention are provided as sets. Thus, a three zinc finger polypeptide comprising fingers F1, F2 and F3 may be presented in a set of

two libraries, each library comprising a two zinc finger polypeptide. A first library is composed of polypeptides consisting essentially of F1 and F2, whilst a second library is composed of polypeptides consisting essentially of F2 and F3. The randomisation in each library includes the overlap between F1 and F2, and F2 and F3 respectively.

5

Preferably, each library will comprise randomisation at at least position 6 of a first finger and position 2 of a second finger. Since these residues contact the same base pair on a double stranded nucleic acid target, it is advantageous that they be varied together.

10

In the case of a three zinc finger polypeptide, the first library will be randomised in fingers F1 and F2, whilst the second is randomised in F2 and F3. Polypeptides may be recombined, post-selection, in the F2 sequence to create a single polypeptide containing F1, F2 and F3. This polypeptide will have been selected taking into account

15

Advantageously, a greater number of position may be varied in each zinc finger. Preferably, residues selected from positions -1, 1, 2, 3 5 and 6 are varied in a first zinc finger and positions -1, 1, 2 and 3 in a second. In a companion library, positions 3, 5 and 6 may be varied in the second finger, and positions -1, 1, 2 and 3 in a third finger. In the final finger (in the case of a three finger protein this will be the third finger), residues 5 and 6 may also be varied.

20

In order that the libraries may be recombined after selection, the polypeptides are preferably designed to include a suitable restriction site in the nucleic acid encoding the zinc finger shared by two libraries. The position of the cleavage site will dictate the precise site of the variations made in the shared zinc finger in each library. Thus, in a set of two libraries encoding a three zinc finger protein, if the cleavage site is between positions 3 and 5 of the α -helix, positions 3 and 5 may be randomised in a first library and positions 5 and 6 in a second.

30

Although it is preferred that residues for randomisation or variation be selected from positions -1, 1, 2, 3, 5 and 6, further residues may also be randomised. For example, the randomisation of position 8 may be advantageous. Moreover, it is envisaged that fewer than all of the given positions are randomised.

5

In a preferred embodiment, a two-library system for selection of a three-finger protein is varied at F1 positions -1, 2, 3, 5, and 6 and F2 positions -1, 1, 2 and 3 in the first library. The second library is varied at F2 positions 3 and 6 and F3 positions -1, 1, 2, 3, 5 and 6. In this case, the cleavage and recombination point will be between residues 3 and 5, preferably between residues 4 and 5, of the α -helix of F2.

10

Subsequent to the recombination event, recombined polypeptide-encoding nucleic acids may be expressed in suitable expression systems, or cloned into Fd phage for further selection.

15

In a preferred aspect of the present invention, the libraries of the invention are not truly randomised at the selected positions, but only partially randomised so that certain but not all amino acids are encoded. This strategy may be used for two purposes.

20

In a first embodiment, variation is restricted to those amino acids which are known to be capable of directing sequence-specific binding of nucleic acid target sequences when incorporated at a given position in the α -helix of a zinc finger. It is known that certain amino acids are not suitable for incorporation at certain positions, irrespective of target sequence. These amino acids are avoided.

25

In a second embodiment, variation is restricted to those amino acids which are known to be capable of directing sequence-specific binding of nucleic acid target sequences when incorporated at a given position in the α -helix of a zinc finger, and variation is directed to specify those residues which are known to favour binding to a specific target sequence at any given position. Thus, the invention permits the design of dedicated

30

libraries from which polypeptides capable of binding to specific target sequence, or to a series of related target sequences, may be selected.

In the first embodiment, which provides a library system for general application, randomisation is preferably effected at all of the positions indicated above. Preferably, the amino acids selected to appear at each given position are as set forth in Table 1:

Position	Possible Amino Acids
-1	R, Q, H, N, D, A, T
1	S, R, K, N
2	D, A, R, Q, H, K, S, N
3	H, N, S, T, V, A, D
5	I, T, K
6	R, Q, V, A, E, K, N, T

TABLE 1

10

It is not necessary for each finger to be randomised at each of the positions given in table 1. In a preferred embodiment, a library for selecting a three-finger protein is constructed according to the specifications given in Table 2:

15

Library 1		Library 2	
F1:	amino acid	F1:	amino acid
-1	R, Q, H, N, D, A		
2	D, A, R, Q, H, K, S, N		
3	H, N, S, T, V, A, D		
5	I, T		
6	R, Q, V, A, E, K, N, T		
F2			
-1	R, Q, H, N, D, A, T		
1	S, R		
2	D, A, R, Q, H, K, S, N		
3	H, N, S, T, V, A, D	3	H, N, S, T, V, A, D
		6	R, Q, V, A, E, K, N, T
F3			
		-1	R, Q, H, N, D, A, T
		1	R, K, S, N
		2	D, A, R, Q, H, K, S, N
		3	H, N, S, T, V, A, D
		5	K, I, T
		6	R, Q, V, A, E, K, N, T

TABLE 2

In the second embodiment, the identity of each amino acid at any particular position is selected according to zinc finger recognition rules as provided herein. In a preferred aspect, therefore, the invention provides a method for preparing a nucleic acid binding protein of the Cys2-His2 zinc finger class capable of binding to a nucleic acid quadruplet in a target nucleic acid sequence, wherein binding to each base of the

quadruplet by an α -helical zinc finger nucleic acid binding motif in the protein is determined as follows:

- a) if base 4 in the quadruplet is G, then position +6 in the α -helix is Arg or Lys;
- 5 b) if base 4 in the quadruplet is A, then position +6 in the α -helix is Glu, Asn or Val;
- c) if base 4 in the quadruplet is T, then position +6 in the α -helix is Ser, Thr, Val or Lys;
- d) if base 4 in the quadruplet is C, then position +6 in the α -helix is Ser, Thr, Val, Ala, Glu or Asn;
- 10 e) if base 3 in the quadruplet is G, then position +3 in the α -helix is His;
- f) if base 3 in the quadruplet is A, then position +3 in the α -helix is Asn;
- g) if base 3 in the quadruplet is T, then position +3 in the α -helix is Ala, Ser or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;
- h) if base 3 in the quadruplet is C, then position +3 in the α -helix is Ser, Asp, Glu,
- 15 Leu, Thr or Val;
- i) if base 2 in the quadruplet is G, then position -1 in the α -helix is Arg;
- j) if base 2 in the quadruplet is A, then position -1 in the α -helix is Gln;
- k) if base 2 in the quadruplet is T, then position -1 in the α -helix is His or Thr;
- l) if base 2 in the quadruplet is C, then position -1 in the α -helix is Asp or His.
- 20 m) if base 1 in the quadruplet is G, then position +2 is Glu;
- n) if base 1 in the quadruplet is A, then position +2 Arg or Gln;
- o) if base 1 in the quadruplet is C, then position +2 is Asn, Gln, Arg, His or Lys;
- p) if base 1 in the quadruplet is T, then position +2 is Ser or Thr.
- 25 The foregoing represents a set of rules which permits the design of a zinc finger binding protein specific for any given nucleic acid sequence. A novel finding related thereto is that position +2 in the helix is responsible for determining the binding to base 1 of the quadruplet. In doing so, it cooperates synergistically with position +6, which determines binding at base 4 in the quadruplet, bases 1 and 4 being overlapping
- 30 in adjacent quadruplets.

Although zinc finger polypeptides are considered to bind to overlapping quadruplet sequences, the method of the present invention allows polypeptides to be designed to bind to target sequences which are not multiples of overlapping quadruplets. For example, a zinc finger polypeptide may be designed to bind to a palindromic target sequence. Such sequences are commonly found as, for example, restriction enzyme target sequences.

Preferably, creation of zinc fingers which bind to fewer than three nucleotides is achieved by specifying, in the zinc finger, amino acids which are unable to support H-bonding with the nucleic acid in the relevant position.

Advantageously, this is achieved by substituting Gly at position -1 (to eliminate a contact with base 2) and/or Ala at positions +3 and/or +6 (to eliminate contacts at the 3rd or 4th base respectively).

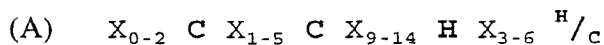
Preferably, the contact with the final (3') base in the target sequence should be strengthened, if necessary, by substituting a residue at the relevant position which is capable of making a direct contact with the phosphate backbone of the nucleic acid.

These and other considerations may be incorporated in a library set in accordance with the invention.

A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA) 85:99-102; Lee *et al.*, (1989) Science 245:635-637; see International patent applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

As used herein, "nucleic acid" refers to both RNA and DNA, constructed from natural nucleic acid bases or synthetic bases, or mixtures thereof. Preferably, however, the binding proteins of the invention are DNA binding proteins.

In general, a preferred zinc finger framework has the structure:

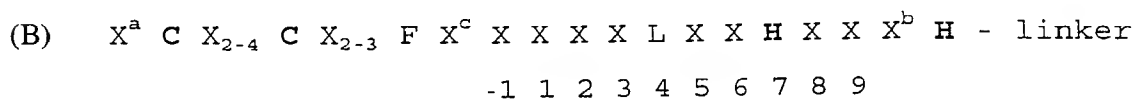


5

where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:

10



15 wherein X (including X^a , X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α -helix.

20 Modifications to this representation may occur or be effected without necessarily abolishing zinc finger function, by insertion, mutation or deletion of amino acids. For example it is known that the second His residue may be replaced by Cys (Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by

25 any aromatic other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for the zinc finger are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α -helix co-ordinated by a zinc atom which contacts four Cys or His residues, does not alter. As used herein,

30 structures (A) and (B) above are taken as an exemplary structure representing all zinc finger structures of the Cys2-His2 type.

Preferably, X^a is $F/Y-X$ or $P-F/Y-X$. In this context, X is any amino acid. Preferably, in this context X is E, K, T or S. Less preferred but also envisaged are Q, V, A and P. The remaining amino acids remain possible.

5

Preferably, X_{2-4} consists of two amino acids rather than four. The first of these amino acids may be any amino acid, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these amino acids is preferably E, although any amino acid may be used.

10

Preferably, X^b is T or I.

Preferably, X^c is S or T.

- 15 Preferably, X_{2-3} is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

Preferably, the linker is T-G-E-K or T-G-E-K-P.

- 20 As set out above, the major binding interactions occur with amino acids -1, +2, +3 and +6. Amino acids +4 and +7 are largely invariant. The remaining amino acids may be essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr.

25

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger nucleic acid binding motif which will bind specifically to a given nucleic acid quadruplet.

- 30 The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid

allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins which are capable of binding to every defined target nucleic acid quadruplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target quadruplet.

In a further aspect of the present invention, there is provided a method for preparing a nucleic acid binding protein of the Cys2-His2 zinc finger class capable of binding to a target nucleic acid sequence, comprising the steps of:

a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and

b) mutating one or more of positions -1, +2, +3 and +6 of the finger as required according to the rules set forth above.

In general, naturally occurring zinc fingers may be selected from those fingers for which the nucleic acid binding specificity is known. For example, these may be the fingers for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

Consensus zinc finger structures may be prepared by comparing the sequences of known zinc fingers, irrespective of whether their binding domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

- 10 The consensuses are derived from the consensus provided by Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523 and from Jacobs, (1993) PhD thesis, University of Cambridge, UK. In both cases, the linker sequences described above for joining two zinc finger motifs together, namely TGEK or TGEKP can be formed on the ends of the consensus. Thus, a P may be removed where necessary, or, in the case of the
15 consensus terminating T G, E K (P) can be added.

When the nucleic acid specificity of the model finger selected is known, the mutation of the finger in order to modify its specificity to bind to the target nucleic acid may be directed to residues known to affect binding to bases at which the natural and desired
20 targets differ. Otherwise, mutation of the model fingers should be concentrated upon residues -1, +2, +3 and +6 as provided for in the foregoing rules.

In order to produce a binding protein having improved binding, moreover, the rules provided by the present invention may be supplemented by physical or virtual
25 modelling of the protein/nucleic acid interface in order to assist in residue selection.

Zinc finger binding motifs designed according to the invention may be combined into nucleic acid binding proteins having a multiplicity of zinc fingers. Preferably, the proteins have at least two zinc fingers. In nature, zinc finger binding proteins
30 commonly have at least three zinc fingers, although two-zinc finger proteins such as Tramtrack are known. The presence of at least three zinc fingers is preferred. Binding

proteins may be constructed by joining the required fingers end to end, N-terminus to C-terminus. Preferably, this is effected by joining together the relevant nucleic acid coding sequences encoding the zinc fingers to produce a composite coding sequence encoding the entire binding protein. The invention therefore provides a method for producing a nucleic acid binding protein as defined above, wherein the nucleic acid binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

- a) preparing a nucleic acid coding sequence encoding two or more zinc finger binding motifs as defined above, placed N-terminus to C-terminus;
- b) inserting the nucleic acid sequence into a suitable expression vector; and
- c) expressing the nucleic acid sequence in a host organism in order to obtain the nucleic acid binding protein.

- 15 A "leader" peptide may be added to the N-terminal finger. Preferably, the leader peptide is MAEEKP.

The nucleic acid encoding the nucleic acid binding protein according to the invention can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the person of ordinary skill in the art. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether it is to be used for DNA amplification or for nucleic acid expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding the nucleic acid binding protein is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise nucleic acid binding protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript® vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up nucleic acid binding protein nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes the nucleic acid binding protein. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to nucleic acid binding protein encoding nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably

linked to DNA encoding the nucleic acid binding protein by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native nucleic acid binding protein promoter sequence and many heterologous promoters may be used to direct
5 amplification and/or expression of nucleic acid binding protein encoding DNA.

Promoters suitable for use with prokaryotic hosts include, for example, the β -lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (Trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have
10 been published, thereby enabling the skilled worker operably to ligate them to DNA encoding nucleic acid binding protein, using linkers or adapters to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding the nucleic acid binding protein.

15

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phagex or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase
20 (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the λ -lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible lac UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be
25 introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL) , vectors containing the trc promoters such as pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) or vectors containing the tac promoter such as pKK223-3
30 (Pharmacia Biotech) or PMAL (New England Biolabs, MA, USA).

Moreover, the nucleic acid binding protein gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space,
5 or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a *Saccharomyces cerevisiae* gene. Thus, the promoter of the TRP1 gene, the ADHI or
10 ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the α - or α -factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-
15 phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid
20 promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PH05 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the
25 PH05 gene.

Nucleic acid binding protein gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus,
30 cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a

ribosomal protein promoter, and from the promoter normally associated with nucleic acid binding protein sequence, provided such promoters are compatible with the host cell systems.

- 5 Transcription of a DNA encoding nucleic acid binding protein by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the
10 late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to nucleic acid binding protein DNA, but is preferably located at a site 5' from the promoter.

- Advantageously, a eukaryotic expression vector encoding a nucleic acid binding protein
15 according to the invention may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the nucleic acid binding protein gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, or in
20 transgenic animals.

- Eukaryotic vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These
25 regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding nucleic acid binding protein.

- An expression vector includes any vector capable of expressing nucleic acid binding protein nucleic acids that are operatively linked with regulatory sequences, such as
30 promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage,

recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which
5 integrate into the host cell genome. For example, DNAs encoding nucleic acid binding protein may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

10 Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding nucleic acid binding protein in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of nucleic
15 acid binding protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying nucleic acid binding protein mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation
20 techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing nucleic acid binding protein
25 expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those
30 skilled in the art will readily envisage how these methods may be modified, if desired.

In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryote, yeast and higher eukaryote cells may be used for replicating DNA and producing the nucleic acid binding protein. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E. coli*, e.g. *E. coli* K-12 strains, DH5a and HB101, or Bacilli. Further hosts suitable for the nucleic acid binding protein encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. *Saccharomyces cerevisiae*. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells including human cells or nucleated cells from other multicellular organisms. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of the nucleic acid binding protein-encoding nucleic acid to form the nucleic acid binding protein. The precise amounts of DNA encoding the nucleic acid binding protein may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient

media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

- 10 Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press).

15

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby the nucleic acid binding protein encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

20

- Nucleic acid binding proteins according to the invention may be employed in a wide variety of applications, including diagnostics and as research tools. Advantageously, they may be employed as diagnostic tools for identifying the presence of nucleic acid molecules in a complex mixture. nucleic acid binding molecules according to the invention can differentiate single base pair changes in target nucleic acid molecules.

25

Accordingly, the invention provides a method for determining the presence of a target nucleic acid molecule, comprising the steps of:

30

- a) preparing a nucleic acid binding protein by the method set forth above which is specific for the target nucleic acid molecule;
- b) exposing a test system comprising the target nucleic acid molecule to the nucleic acid binding protein under conditions which promote binding, and removing any nucleic acid binding protein which remains unbound;
- c) detecting the presence of the nucleic acid binding protein in the test system.

In a preferred embodiment, the nucleic acid binding molecules of the invention can be incorporated into an ELISA assay. For example, phage displaying the molecules of the invention can be used to detect the presence of the target nucleic acid, and visualised using enzyme-linked anti-phage antibodies.

Further improvements to the use of zinc finger phage for diagnosis can be made, for example, by co-expressing a marker protein fused to the minor coat protein (gVIII) of bacteriophage. Since detection with an anti-phage antibody would then be obsolete, the time and cost of each diagnosis would be further reduced. Depending on the requirements, suitable markers for display might include the fluorescent proteins (A. B. Cubitt, *et al.*, (1995) *Trends Biochem Sci.* **20**, 448-455; T. T. Yang, *et al.*, (1996) *Gene* **173**, 19-23), or an enzyme such as alkaline phosphatase which has been previously displayed on gIII (J. McCafferty, R. H. Jackson, D. J. Chiswell, (1991) *Protein Engineering* **4**, 955-961). Labelling different types of diagnostic phage with distinct markers would allow multiplex screening of a single nucleic acid sample. Nevertheless, even in the absence of such refinements, the basic ELISA technique is reliable, fast, simple and particularly inexpensive. Moreover it requires no specialised apparatus, nor does it employ hazardous reagents such as radioactive isotopes, making it amenable to routine use in the clinic. The major advantage of the protocol is that it obviates the requirement for gel electrophoresis, and so opens the way to automated nucleic acid diagnosis.

The invention provides nucleic acid binding proteins which can be engineered with exquisite specificity. The invention lends itself, therefore, to the design of any

molecule of which specific nucleic acid binding is required. For example, the proteins according to the invention may be employed in the manufacture of chimeric restriction enzymes, in which a nucleic acid cleaving domain is fused to a nucleic acid binding domain comprising a zinc finger as described herein.

5

Moreover, the invention provides therapeutic agents and methods of therapy involving use of nucleic acid binding proteins as described herein. In particular, the invention provides the use of polypeptide fusions comprising an integrase, such as a viral integrase, and a nucleic acid binding protein according to the invention to target nucleic acid sequences *in vivo* (Bushman, (1994) PNAS (USA) 91:9233-9237). In gene therapy applications, the method may be applied to the delivery of functional genes into defective genes, or the delivery of nonsense nucleic acid in order to disrupt undesired nucleic acid. Alternatively, genes may be delivered to known, repetitive stretches of nucleic acid, such as centromeres, together with an activating sequence such as an LCR. This would represent a route to the safe and predictable incorporation of nucleic acid into the genome.

10

15

20

25

In conventional therapeutic applications, nucleic acid binding proteins according to the invention may be used to specifically knock out cell having mutant vital proteins. For example, if cells with mutant ras are targeted, they will be destroyed because ras is essential to cellular survival. Alternatively, the action of transcription factors may be modulated, preferably reduced, by administering to the cell agents which bind to the binding site specific for the transcription factor. For example, the activity of HIV tat may be reduced by binding proteins specific for HIV TAR.

Moreover, binding proteins according to the invention may be coupled to toxic molecules, such as nucleases, which are capable of causing irreversible nucleic acid damage and cell death. Such agents are capable of selectively destroying cells which comprise a mutation in their endogenous nucleic acid.

30

Nucleic acid binding proteins and derivatives thereof as set forth above may also be applied to the treatment of infections and the like in the form of organism-specific antibiotic or antiviral drugs. In such applications, the binding proteins may be coupled to a nuclease or other nuclear toxin and targeted specifically to the nucleic acids of microorganisms.

The invention likewise relates to pharmaceutical preparations which contain the compounds according to the invention or pharmaceutically acceptable salts thereof as active ingredients, and to processes for their preparation.

The pharmaceutical preparations according to the invention which contain the compound according to the invention or pharmaceutically acceptable salts thereof are those for enteral, such as oral, furthermore rectal, and parenteral administration to (a) warm-blooded animal(s), the pharmacological active ingredient being present on its own or together with a pharmaceutically acceptable carrier. The daily dose of the active ingredient depends on the age and the individual condition and also on the manner of administration.

The novel pharmaceutical preparations contain, for example, from about 10 % to about 80 %, preferably from about 20 % to about 60 %, of the active ingredient. Pharmaceutical preparations according to the invention for enteral or parenteral administration are, for example, those in unit dose forms, such as sugar-coated tablets, tablets, capsules or suppositories, and furthermore ampoules. These are prepared in a manner known per se, for example by means of conventional mixing, granulating, sugar-coating, dissolving or lyophilising processes. Thus, pharmaceutical preparations for oral use can be obtained by combining the active ingredient with solid carriers, if desired granulating a mixture obtained, and processing the mixture or granules, if desired or necessary, after addition of suitable excipients to give tablets or sugar-coated tablet cores.

Suitable carriers are, in particular, fillers, such as sugars, for example lactose, sucrose, mannitol or sorbitol, cellulose preparations and/or calcium phosphates, for example tricalcium phosphate or calcium hydrogen phosphate, furthermore binders, such as starch paste, using, for example, corn, wheat, rice or potato starch, gelatin, tragacanth, methylcellulose and/or polyvinylpyrrolidone, if desired, disintegrants, such as the abovementioned starches, furthermore carboxymethyl starch, crosslinked polyvinylpyrrolidone, agar, alginic acid or a salt thereof, such as sodium alginate; auxiliaries are primarily glidants, flow-regulators and lubricants, for example silicic acid, talc, stearic acid or salts thereof, such as magnesium or calcium stearate, and/or polyethylene glycol. Sugar-coated tablet cores are provided with suitable coatings which, if desired, are resistant to gastric juice, using, inter alia, concentrated sugar solutions which, if desired, contain gum arabic, talc, polyvinylpyrrolidone, polyethylene glycol and/or titanium dioxide, coating solutions in suitable organic solvents or solvent mixtures or, for the preparation of gastric juice-resistant coatings, solutions of suitable cellulose preparations, such as acetylcellulose phthalate or hydroxypropylmethylcellulose phthalate. Colorants or pigments, for example to identify or to indicate different doses of active ingredient, may be added to the tablets or sugar-coated tablet coatings.

Other orally utilisable pharmaceutical preparations are hard gelatin capsules, and also soft closed capsules made of gelatin and a plasticiser, such as glycerol or sorbitol. The hard gelatin capsules may contain the active ingredient in the form of granules, for example in a mixture with fillers, such as lactose, binders, such as starches, and/or lubricants, such as talc or magnesium stearate, and, if desired, stabilisers. In soft capsules, the active ingredient is preferably dissolved or suspended in suitable liquids, such as fatty oils, paraffin oil or liquid polyethylene glycols, it also being possible to add stabilisers.

Suitable rectally utilisable pharmaceutical preparations are, for example, suppositories, which consist of a combination of the active ingredient with a suppository base. Suitable suppository bases are, for example, natural or synthetic triglycerides, paraffin

hydrocarbons, polyethylene glycols or higher alkanols. Furthermore, gelatin rectal capsules which contain a combination of the active ingredient with a base substance may also be used. Suitable base substances are, for example, liquid triglycerides, polyethylene glycols or paraffin hydrocarbons.

5

Suitable preparations for parenteral administration are primarily aqueous solutions of an active ingredient in water-soluble form, for example a water-soluble salt, and furthermore suspensions of the active ingredient, such as appropriate oily injection suspensions, using suitable lipophilic solvents or vehicles, such as fatty oils, for example sesame oil, or synthetic fatty acid esters, for example ethyl oleate or triglycerides, or aqueous injection suspensions which contain viscosity-increasing substances, for example sodium carboxymethylcellulose, sorbitol and/or dextran, and, if necessary, also stabilisers.

10

15 The dose of the active ingredient depends on the warm-blooded animal species, the age and the individual condition and on the manner of administration. In the normal case, an approximate daily dose of about 10 mg to about 250 mg is to be estimated in the case of oral administration for a patient weighing approximately 75 kg .

20

The invention is described below, for the purpose of illustration only, in the following examples.

Example 1

Determination of binding site preferences in zinc fingers

25

Design Of Zinc Finger Phage Display Libraries

Zinc finger-DNA recognition at the interface between adjacent DNA subsites is studied using a zinc finger phage display library. This library is based on the three-finger DNA-binding domain of Zif268, but contains randomisations of amino acids from finger 2 (F2) and finger 3 (F3), at residue positions which could form a network of contacts across the interface of their DNA subsites. The detailed design of the library

30

is shown in Figure 1c, together with the generic DNA binding site used in selections. Briefly, the library contains randomisations at F2 residue position 6 (hereafter denoted F2[+6]) and F3 residue positions -1, +1, +2 and +3 (hereafter denoted F3[-1], F3[+2], etc.).

5

Library selections are carried out using DNA binding sites that resembled the Zif268 operator, but which contained systematic combinations of bases in the DNA doublet which forms the base-step between the DNA subsites of F2 and F3. DNA binding sites are of the generic form 5'-GNX-XCG-GCG-3', where X-X denotes a given
10 combination of the bases at the interface between the DNA subsites, and N denotes that the four bases are equally represented at DNA position 3. Thus the interaction between F3[+3] and nucleotide position 3N is allowed complete freedom in this experiment. This feature of the library allows selection of a large family (or database) of related zinc fingers that bind a given combination of bases at nucleotide positions 4X and 5X,
15 but which are non-identical owing to different interactions with the middle base in the nominal triplet subsite of F3.

The first library to be constructed, LIB-A, contains randomisations at F2 residue position 6 and F3 residue positions -1, 1, 2 and 3 (see Figure 2), and is sorted using the
20 DNA sequence 5'-GNX-XCG-GCG-3', where X-X denotes a known combination of the two bases at DNA positions 4X and 5X, and N denotes an equal probability of any of the four bases at DNA position 3. The second library, LIB-B, contains randomisations at F2 residue position 6 and F3 residue positions -1 and 2, and is sorted using the DNA sequence 5'-GCX-XCG-GCG-3', where X-X denotes a known combination of the two
25 bases at DNA positions 4X and 5X.

The genes for the two different zinc finger phage display libraries are assembled from four synthetic DNA oligonucleotides by directional end-to-end ligation using three short complementary DNA linkers. The oligonucleotides contain selectively randomised
30 codons (of sequence NNS; N = A/C/G/T, S = G/C) in the appropriate amino acid positions of fingers 2 and 3. The constructs are amplified by PCR using primers

containing *Not I* and *Sfi I* restriction sites, digested with the above endonucleases to produce cloning overhangs, and ligated into phage vector Fd-Tet-SN. Electrocompetent *E. coli* TG 1 cells are transformed with the recombinant vector and plated onto TYE medium (1.5% agar, 1% Bacto tryptone, 0.5% Bacto yeast extract, 0.8% NaCl) containing 15 µg/ml tetracycline.

Allowing this freedom to some protein-DNA interactions that are not being studied is a useful strategy towards increasing the diversity of clones which can be obtained from any one selection experiment. However, at the same time, it is important to limit the number of contacts that are allowed contextual freedom at any one time, otherwise there is a danger that a subset of particularly strong intermolecular interactions will dominate the selections. Anticipating this eventuality, a smaller sublibrary is also created that contains randomised residues only in positions F2[+6] and F3[-1 and +2], and therefore does not allow for contextual freedom in selections. Clones selected from this library are marked with an asterisk when they are discussed herein.

Experimental Strategy

Phage selections from the two zinc finger libraries are performed separately in order to determine the diversity of DNA sequences which can be bound specifically by members of each library. Sixteen selections are performed on each library, using the different DNA binding sites that correspond to all 16 possible combinations of bases at nucleotide positions 4X and 5X. The DNA binding site used to select specifically binding phage is immobilised on a solid surface, while a 10-fold excess of each of the other 15 DNA sites is present in solution as a specific competitor.

Phage Selections

Tetracycline resistant colonies are transferred from plates into 2xTY medium (16g/litre Bacto tryptone, 10g/litre Bacto yeast extract, 5g/litre NaCl) containing 50µM ZnCl₂ and 15 µg/ml tetracycline, and cultured overnight at 30°C in a shaking incubator. Cleared culture supernatant containing phage particles is obtained by centrifuging at 300g for 5 minutes.

Biotinylated DNA target sites (1pmol) are bound to streptavidin-coated tubes (Boehringer Mannheim). Phage supernatant solutions are diluted 1:10 in PBS selection buffer (PBS containing 50µM ZnCl₂, 2% Marvel, 1% Tween, 20µg/ml sonicated salmon sperm DNA, 10 pmol/ml of each of the 15 other possible unbiotinylated DNA sites), and 1 ml is applied to each tube for 1 hour at 20°C. After this time, the tubes are emptied and washed 20 times with PBS containing 50µM ZnCl₂, 2% Marvel and 1% Tween. Retained phage are eluted in 0.1ml 0.1M triethylamine and neutralised with an equal volume of 1M Tris (pH 7.4). Logarithmic-phase *E. coli* TG 1 (0.5ml) are infected with eluted phage (50µl), and used to prepare phage supernatants for subsequent rounds of selection. After 3 rounds of selection, *E. coli* infected with selected phage are plated, individual colonies are picked and used to grow phage for binding site signature assays and DNA sequencing.

After three rounds of phage selection against a particular DNA binding site, individual zinc finger clones are recovered, and the DNA binding specificity of each clone is determined by the binding site signature method. This involves screening each zinc finger phage for binding to eight different libraries of the DNA binding site, designed such that each library contains one fixed base and one randomised base at either of positions 4X and 5X (i.e. libraries GN, AN, TN, CN, and NG, NA, NT, NC). Thus each of the 16 DNA binding sites used in selection experiments is specified by a unique combination of two libraries - for example, the DNA binding site containing 4G5G is present in only two of the eight libraries in which the relevant doublet had one nucleotide randomised and the other nucleotide fixed as guanine, i.e. libraries 4G5N and 4N5G. The eight DNA libraries used in binding site signatures are arrayed across a microtitre plate and zinc finger phage binding is detected by phage ELISA. The pattern of binding to the eight DNA libraries reveals the DNA sequence specificity (or preference) of each phage clone, and only those clones found to be relatively specific are subsequently sequenced to reveal the identity of the amino acids present in the randomised zinc finger residue positions.

- Procedures are as described previously (Choo, Y. & Klug, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11163-11167; Choo, Y. & Klug, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11168-11172). Briefly, 5'-biotinylated positionally randomised oligonucleotide libraries, containing Zif268 operator variants, are synthesised by primer extension as described. DNA libraries (0.4 pmol/well for LIB-A and 1.2 pmol/well for LIB-B) are added to streptavidin-coated ELISA wells (Boehringer-Mannheim) in PBS containing 50µM ZnCl₂ (PBS/Zn). Phage solution (overnight bacterial culture supernatant diluted 1:10 in PBS/Zn containing 2% Marvel, 1% Tween and 20µg/ml sonicated salmon sperm DNA) are applied to each well (50µl/well). Binding is allowed to proceed for one hour at 20°C. Unbound phage are removed by washing 6 times with PBS/Zn containing 1% Tween, then 3 times with PBS/Zn. Bound phage are detected by ELISA using horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech) and the colourimetric signal quantitated using SOFFMAX 2.32 (Molecular Devices).
- The coding sequence of individual zinc finger clones is amplified by PCR using external primers complementary to phage sequence. These PCR products are sequenced manually using Thermo Sequenase cycle sequencing (Amersham Life Science).
- Analysis Of Phage-Selected Zinc Fingers*
- Figure 3 shows the binding site signatures of relatively sequence-specific zinc finger phages selected from both libraries, using the 16 different DNA doublets which form the base-step between the DNA subsites of fingers 2 and 3. The results show that zinc finger clones are selected which bind specifically to almost all subsites, including those triplets in which the 5' position (nucleotide 5X in the model system) is fixed as a base other than guanine. Overall, the selections show that any of the four bases can be bound specifically in both the 5' and 3' positions of a nominal triplet subsite. The results are summarised in Figure 4.
- Selections from the smaller sub-library yield fingers that can bind specifically to only 8 of the 16 doublets, whereas members of the larger library yield fingers that recognise

15 out of the 16 doublets. It is not known whether this difference in efficacy originates from the inclusion of more randomised positions in the larger library, or the conformational flexibility afforded by the contextual freedom designed into the larger library, or both. The only base-step that does not yield specific zinc fingers is 4G5A.

- 5 This dinucleotide may induce an unfavourable DNA deformation in the context of the DNA binding sites used for selection.

Example 2

Determination of +2 specificity for position 1

10

The amino acid present in α -helical position 2 of a zinc finger can help determine the specificity for the base-pair at the interface of two overlapping DNA quadruplet subsites (see Figure 1B; position 5/5', corresponding to position 1 or 4 of the quadruplet as discussed above). An Asp residue present in F3[+2] of wild-type Zif268
15 has been shown to play a role in DNA recognition, and further examples are generated by the current phage display experiments (See Example 1 for details, and Figure 5A).

The experimental protocol followed is that of Example 1. Figure 5A shows an example of related zinc finger clones showing the effect of α -helical position 2 on DNA-binding
20 specificity. In this case, position 6 of finger 2 is invariant (Asn) and the change in case specificity in the zinc finger in order to select for contact to this base is dictated by position +2 in finger 3.

This family of zinc fingers is derived from selections using DNA binding sites
25 containing 4T5A or 4T5C subsite interfaces. The base preference for the 5X- 5'X base-pair is determined by the amino acid present at F3[+2], probably by the formation of cross-strand contacts.

Figure 5B shows examples of correlations between certain amino acids selected at
30 F3[+2] and the identity of the base present at position 5'X. Selections reveal the possibility of DNA contacts from five amino acids (Asn, Gln, Arg, Lys and His) which

are all capable of donating a H-bond to the exocyclic oxygen atom of either guanine (O₆) or thymine (O₄) in nucleotide position 5'X. The clones isolated with these amino acids at F3[+2] are listed in this diagram together with the binding site signature showing the base-preference at position 5'X. Overall, Ser dominated the selections with an occurrence of 38%, in accord with its presence in position 2 in over half of all known zinc fingers. Threonine, Ala and Gly occurred frequently in the selections (15%, 15% and 9% respectively) but did not show any discernible patterns of discrimination. Certain amino acids (Cys, Asp, Phe, Ile, Leu, Met, Pro, Val and Trp) are never selected in position 2. Their ability to bind in certain situations is however not to be excluded.

A small subset of amino acids selected in F3[+2] show significant correlations to the identity of the base-pair in position 5'X (Figure 5B), suggesting that cross-strand interactions between these may be a general mechanism of DNA-recognition. Most of these correlations can be rationalised as pairings between hydrogen bond donors in F3[+2] and guanine or thymine in DNA position 5'X, in accordance with the framework of the Zif268 model. In contrast to amino acids that are never selected in position 2, or amino acids that are selected but which show no significant correlations, the amino acids which consistently appear to play a role in DNA recognition from this position have side chains with multiple hydrogen bonding groups. It is possible that these residues can play a role in base recognition because they achieve greater specificity by participating in buttressing networks.

Example 3

Construction of a General Purpose Library

The binary library system constructed in this example comprises libraries LIB1/2 and LIB2/3 that each encode the three fingers of Zif268 but with some amino acid positions selectively randomised. Instead of adhering to the model of modular zinc fingers, the new libraries contain concerted variations in certain amino acid positions in adjacent zinc fingers. Thus LIB1/2 contains simultaneous variations in F1 positions -1, 2, 3, 5

and 6 and F2 positions -1, 1, 2 and 3. LIB2/3 contains simultaneous variations in F2 positions 3 and 6 and F3 positions -1, 1, 2, 3 and 5, 6. The remaining amino acids in each library are as the WT Zif268 sequence. The two libraries are cloned in Fd phage as GIII fusions according to standard protocols.

5

The amino acids that are allowed at each varied position are as follows:

LIB1/2

F1 pos. -1 = R, Q, H, N, D, A, T;

pos. 2 = D, A, R, Q, H, K, S, N;

10 pos. 3 = H, N, S, T, V, A, D;

pos. 5 = I, T;

pos. 6 = R, Q, V, A, E, K, N, T.

F2 pos. -1 = R, Q, H, N, D, A, T;

pos. 1 = S, R;

15 pos. 2 = D, A, R, Q, H, K, S, N;

pos. 3 = H, N, S, T, V, A, D;

LIB2/3

F2 pos. 3 = H, N, S, T, V, A, D;

20 pos. 6 = R, Q, V, A, E, K, N, T.

F3 pos. -1 = R, Q, H, N, D, A, T;

pos. 1 = R, K, S, N;

pos. 2 = D, A, R, Q, H, K, S, N;

pos. 3 = H, N, S, T, V, A, D;

25 pos. 5 = K, I, T;

pos. 6 = R, Q, V, A, E, K, N, T.

Selections And Recombinations

Selections are performed using the DNA sequence GCG-GMN-OPQ for LIB 1/2 and
 30 the DNA sequence IJK-LMG-GCG for LIB2/3, where the underlined bases are bound
 by the WT Zif268 residues and each of the other letters stands for any given nucleotide.
 The conserved nucleotides of the Zif268 binding site serve to fix the register of the

interaction by binding to the conserved portion of the Zif268 DNA-binding domain. The binary phage display libraries can be mixed so that selections using these two generic sites are performed in a single tube, or the selections can be performed separately. After a number of rounds of selection the two libraries are recombined to
5 produce a chimaeric DNA-binding domain that recognises the sequence IJK-LMN-OPQ.

The recombination reactions are performed by amplifying the selected three-finger domains by PCR and cutting the PCR products using restriction enzyme Ddel. This
10 cuts the genes of both zinc finger libraries at the DNA sequence coding for F2 α -helical positions 4 and 5. The digested products are randomly religated to produce recombinant genes coding for the chimaeric DNA-binding domains (and other products including reconstituted WT Zif268). The chimaeric DNA-binding domains are selectively amplified from the mixture of products by PCR using selective primers that recognise
15 the recombinant F1 and F3 genes, rather than WT genes, and cloned in Fd phage (for more selections) or other vectors (e.g. for expression in *E coli*).

The initial selections from the binary libraries can be pushed to completion, thus allowing the assembly of a single clone by recombination. Alternatively, if the initial
20 selections are less stringent, many candidates will be available for the assembly of various chimaeric domains after recombination. In the latter case, the best recombinant protein can be selected by further rounds of selection on phage.

Claims

1. A zinc finger polypeptide library in which each polypeptide comprises more than one zinc finger which has been at least partially randomised.

5

2. A library according to claim 1 wherein two zinc fingers are at least partially randomised in each polypeptide.

10

3. A library according to claim 1 or claim 2, wherein the randomised zinc fingers are adjacent.

15

4. A set of zinc finger polypeptide libraries which encode overlapping zinc finger polypeptides, each polypeptide comprising more than one zinc finger which has been at least partially randomised, and which polypeptides may be assembled after selection to form a multifinger zinc finger polypeptide.

5. A set according to claim 4, comprising a pair of libraries encoding three-zinc finger polypeptides.

20

6. A library or set of libraries according to any preceding claim, wherein the randomised positions are selected from positions -1, 1, 2, 3, 5 and 6.

7. A library according to any preceding claim, wherein the randomisation of amino acid residues is restricted such that the following amino acids may appear at the given positions:

25

Position	Possible Amino Acids
-1	R, Q, H, N, D, A, T
1	S, R, K, N
2	D, A, R, Q, H, K, S, N
3	H, N, S, T, V, A, D
5	I, T, K

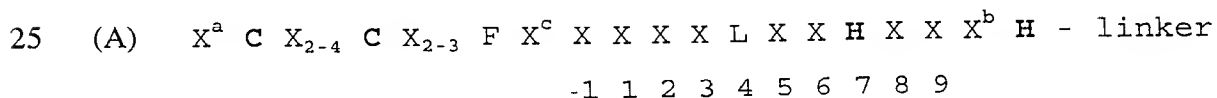
8. A set of two libraries according to claim 7 for selecting a three-finger zinc finger protein, wherein the following amino acids may appear at the given positions:

Library 1		Library 2	
F1:	amino acid	F1:	amino acid
-1	R, Q, H, N, D, A		
2	D, A, R, Q, H, K, S, N		
3	H, N, S, T, V, A, D		
5	I, T		
6	R, Q, V, A, E, K, N, T		
F2			
-1	R, Q, H, N, D, A, T		
1	S, R		
2	D, A, R, Q, H, K, S, N		
3	H, N, S, T, V, A, D	3	H, N, S, T, V, A, D
		6	R, Q, V, A, E, K, N, T
F3			
		-1	R, Q, H, N, D, A, T
		1	R, K, S, N
		2	D, A, R, Q, H, K, S, N
		3	H, N, S, T, V, A, D
		5	K, I, T
		6	R, Q, V, A, E, K, N, T

9. A library according to claim 1, wherein the amino acids at positions -1, 2, 3 and 6 are selected as follows:

- a) if base 4 in the quadruplet is G, then position +6 in the α -helix is Arg or Lys;
- b) if base 4 in the quadruplet is A, then position +6 in the α -helix is Glu, Asn or Val;
- c) if base 4 in the quadruplet is T, then position +6 in the α -helix is Ser, Thr, Val or Lys;
- 5 d) if base 4 in the quadruplet is C, then position +6 in the α -helix is Ser, Thr, Val, Ala, Glu or Asn;
- e) if base 3 in the quadruplet is G, then position +3 in the α -helix is His;
- f) if base 3 in the quadruplet is A, then position +3 in the α -helix is Asn;
- g) if base 3 in the quadruplet is T, then position +3 in the α -helix is Ala, Ser or Val;
- 10 provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;
- h) if base 3 in the quadruplet is C, then position +3 in the α -helix is Ser, Asp, Glu, Leu, Thr or Val;
- i) if base 2 in the quadruplet is G, then position -1 in the α -helix is Arg;
- j) if base 2 in the quadruplet is A, then position -1 in the α -helix is Gln;
- 15 k) if base 2 in the quadruplet is T, then position -1 in the α -helix is His or Thr;
- l) if base 2 in the quadruplet is C, then position -1 in the α -helix is Asp or His.
- m) if base 1 in the quadruplet is G, then position +2 is Glu;
- n) if base 1 in the quadruplet is A, then position +2 Arg or Gln;
- o) if base 1 in the quadruplet is C, then position +2 is Asn, Gln, Arg, His or Lys;
- 20 if base 1 in the quadruplet is T, then position +2 is Ser or Thr.

10. A library according to any preceding claim, wherein each zinc finger has the general primary structure



wherein X (including X^a, X^b and X^c) is any amino acid.

- 30 11. A library according to claim 10 wherein X^a is $F/Y-X$ or $P-F/Y-X$.

12. A library according to claim 10 or claim 11 wherein X_{2-4} is selected from any one of: S-X, E-X, K-X, T-X, P-X and R-X.

13. A library according to any one of claims 10 to 12 wherein X^b is T or I.

5

14. A library according to any one of claims 10 to 13 wherein X_{2-3} is G-K-A, G-K-C, G-K-S, G-K-G, M-R-N or M-R.

15. A library according to any one of claims 10 to 14 wherein the linker is T-G-E-K or T-G-E-K-P.

10

16. A library according to any one of claims 10 to 15 wherein position +9 is R or K.

17. A library according to any one of claims 10 to 16 wherein positions +1, +5 and +8 are not occupied by any one of the hydrophobic amino acids, F, W or Y.

15

18. A library according to claim 17 wherein positions +1, +5 and +8 are occupied by the residues K, T and Q respectively.

20

19. A method for preparing a library of nucleic acid binding proteins of the Cys2-His2 zinc finger class capable of binding to a target nucleic acid sequence, comprising the steps of:

a) selecting a model zinc finger polypeptide from the group consisting of naturally occurring zinc finger polypeptides and consensus zinc finger polypeptides; and

25

b) randomising more than one finger therein according to any one of claims 1 to 9.

20. A method according to claim 19, wherein the model zinc finger is a consensus zinc finger whose structure is selected from the group consisting of the consensus

30

structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

21. A method according to claim 19 wherein the model zinc finger is a naturally occurring zinc finger whose structure is selected from one finger of a protein selected from the group consisting of Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

22. A method according to claim 21 wherein the model zinc finger is finger 2 of Zif 268.

23. A method for determining the presence of a target nucleic acid molecule, comprising the steps of:

- a) preparing a nucleic acid binding protein by the method of any preceding claim which is specific for the target nucleic acid molecule;
- b) exposing a test system comprising the target nucleic acid molecule to the nucleic acid binding protein under conditions which promote binding, and removing any nucleic acid binding protein which remains unbound;
- c) detecting the presence of the nucleic acid binding protein in the test system.

24. A method according to claim 23, wherein the presence of the nucleic acid binding protein in the test system is detected by means of an antibody.

25. A method according to claim 23 or claim 24 wherein the nucleic acid binding protein, in use, is displayed on the surface of a filamentous bacteriophage and the presence of the nucleic acid binding protein is detected by detecting the bacteriophage or a component thereof.

1/6

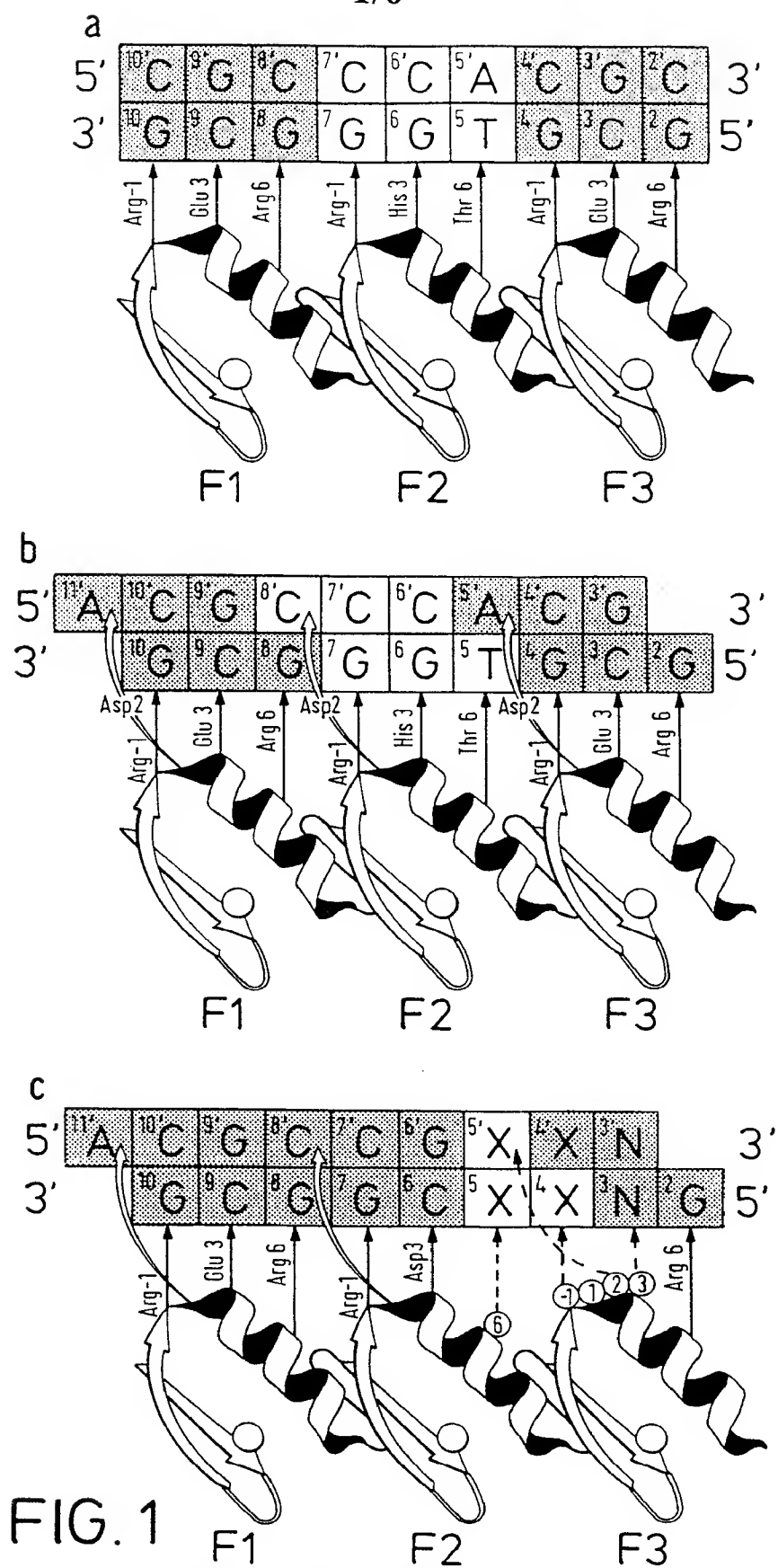


FIG. 1

SUBSTITUTE SHEET (RULE 26)

2/6

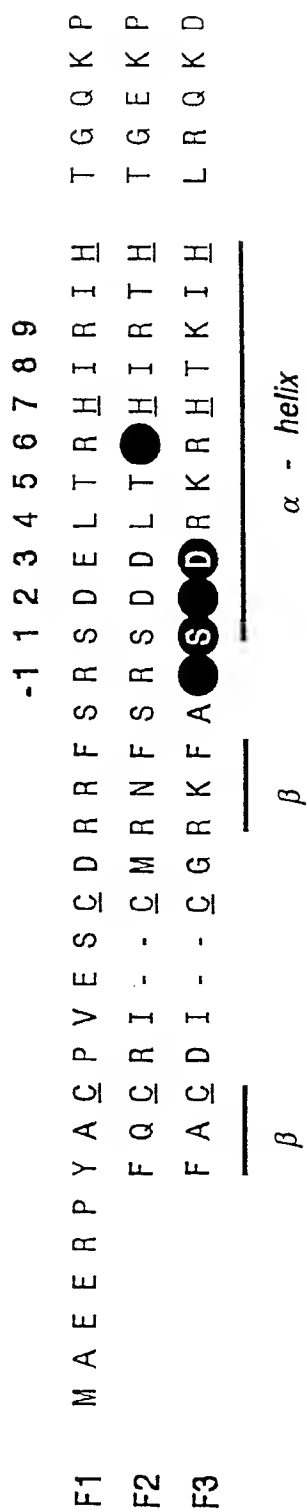


FIG. 2

3/6

BASE AT POSITIONS 4'X AND 5'X OF DNA SITES USED IN SELECTIONS	AMINO ACIDS SELECTED IN FINGERS 2 AND 3	BINDING SITE SIGNATURE	No. OF CLONES SELECTED AND SCREENED	BASES AT POSITIONS 4'X AND 5'X OF DNA SITES USED IN SELECTIONS	AMINO ACIDS SELECTED IN FINGERS 2 AND 3	BINDING SITE SIGNATURE	No. OF CLONES SELECTED AND SCREENED
4'X 5'X	F2 F3	4'X 5'X		4'X 5'X	F2 F3	4'X 5'X	
GG	6 R * R R R R	-1123 RLEY RSED RHTH RSSE RSSA	4 4 1 1 1	CC	6 T E E S T	DPHN HSKS HRQN DRAN DRAN	1 4 1 1 2
AG	R * R R R K	QVTT QSGD QLAT QDAH QRAS QSTS	4 5 2 1 1 1	AC	* Y V T	RSQD QVGH QLAT	3 2 3
TG	R R * R K R R R	SSGD SASA NSGD LVQN TGAS TPSG TQTA TSAA	1 1 1 1 1 1 1 2	TA	N N K * E E V V L	TSQA TSAS LAQT RSRD RLRD HLAT HLTT VGHH	4 2 1 3 4 1 1 1
CG	R R R R	DTSV DAST DASA DTSS	3 2 1 2	CA	Y N E L	HPAT HPAN HHSN DSRA	2 1 9 2
GC	A * N * S S T T T T	RNHD RSTD RSTD SRHS RNST RTST TRYs RAQN	2 3 6 1 1 1 3 1	GT	* T * K * K A T T	KSSD SSSD RSHD RSSY RSSS RKTD	2 1 1 1 1 1
AC	A A * V S S A T	QAAT QGTN TSRD QRGA QSTT TSSS SSST	4 1 4 1 2 1 1	AT	S S T T T T T T K	QIST QIGA QYST QSAS QSQH QTSH QPGH QDST QDST	3 1 1 1 1 1 1 1 2
TC	D N V V V * R A T S S S G	TISN TSTA TSSL TSSI TSNS GSND TTSS TAGS TTSS TSSA LSTT LSST	1 1 4 1 3 1 1 1 1 1 1 1 1	TT	T T T T S S S	TAST TASH TSSV TSSA HHTS HAQT HATT	1 1 1 1 1 1 4
				CT	K T S	DHSS HPST DSSR	5 1 1

RELATIVE SIGNATURE STRENGTH ■ ≤1.00 ■ 0.80 ■ 0.60 ■ 0.40 ■ 0.20 □ 0.00

FIG. 3
SUBSTITUTE SHEET (RULE 26)

4/6

a

AMINO ACID SELECTED AT POSITION -1

BASE PRESENT AT POSITION ⁴ X IN BINDING SIGNATURES	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
G	0	0	0	0	0	0	0	0	1	0	0	0	0	0	16	2	2	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	21	1	1	1	0	0	0
T	0	0	0	0	0	1	7	0	0	4	0	1	0	2	5	3	21	1	0	0
C	0	0	10	0	0	0	7	0	0	2	0	0	0	0	2	0	0	0	0	0

↓ ↓ ↓ ↓ ↓

C T/C T A G(T) T

RECOGNITION PATTERNS

b

AMINO ACID SELECTED AT POSITION 6

BASE PRESENT AT POSITION ⁵ X IN BINDING SIGNATURES	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
G	0	0	0	0	0	0	0	0	4	1	0	0	0	0	21	0	0	0	0	0
A	0	0	0	3	0	0	0	0	1	2	0	3	0	0	0	0	2	3	0	2
T	1	0	0	0	0	1	0	0	5	0	0	0	0	0	0	10	17	3	0	0
C	5	0	1	3	0	1	0	0	0	0	0	3	0	0	2	12	14	6	0	0

↓ ↓ ↓ ↓ ↓ ↓ ↓

C A/C G/T A/C G T/C C(A/T)

RECOGNITION PATTERNS

FIG. 4

5/6

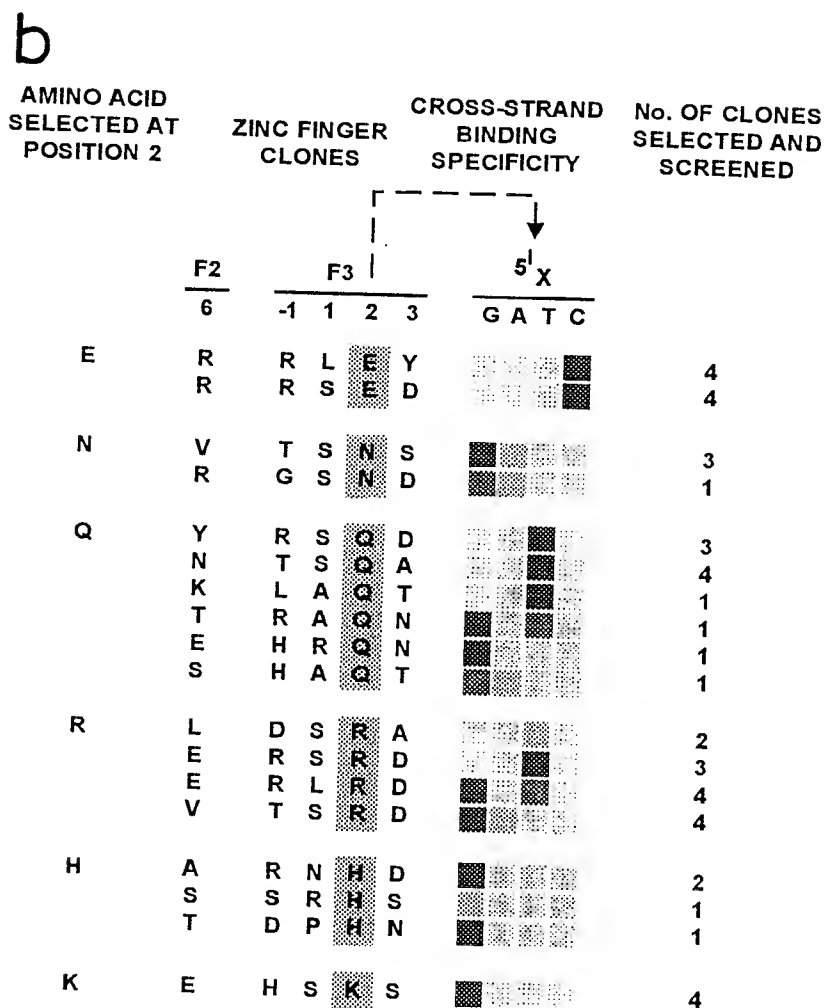
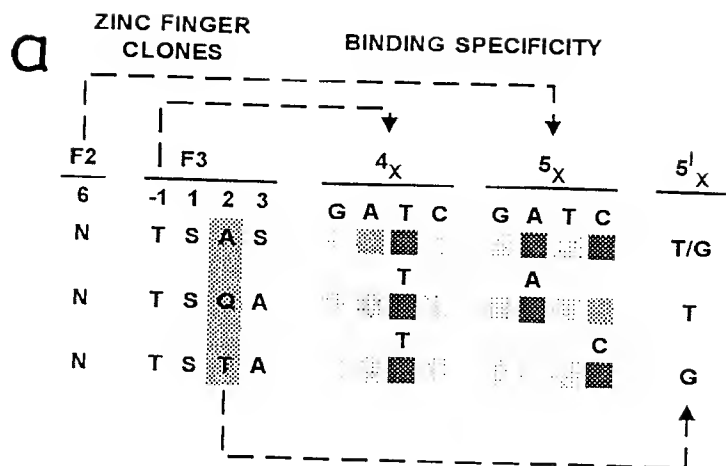


FIG. 5

6 / 6

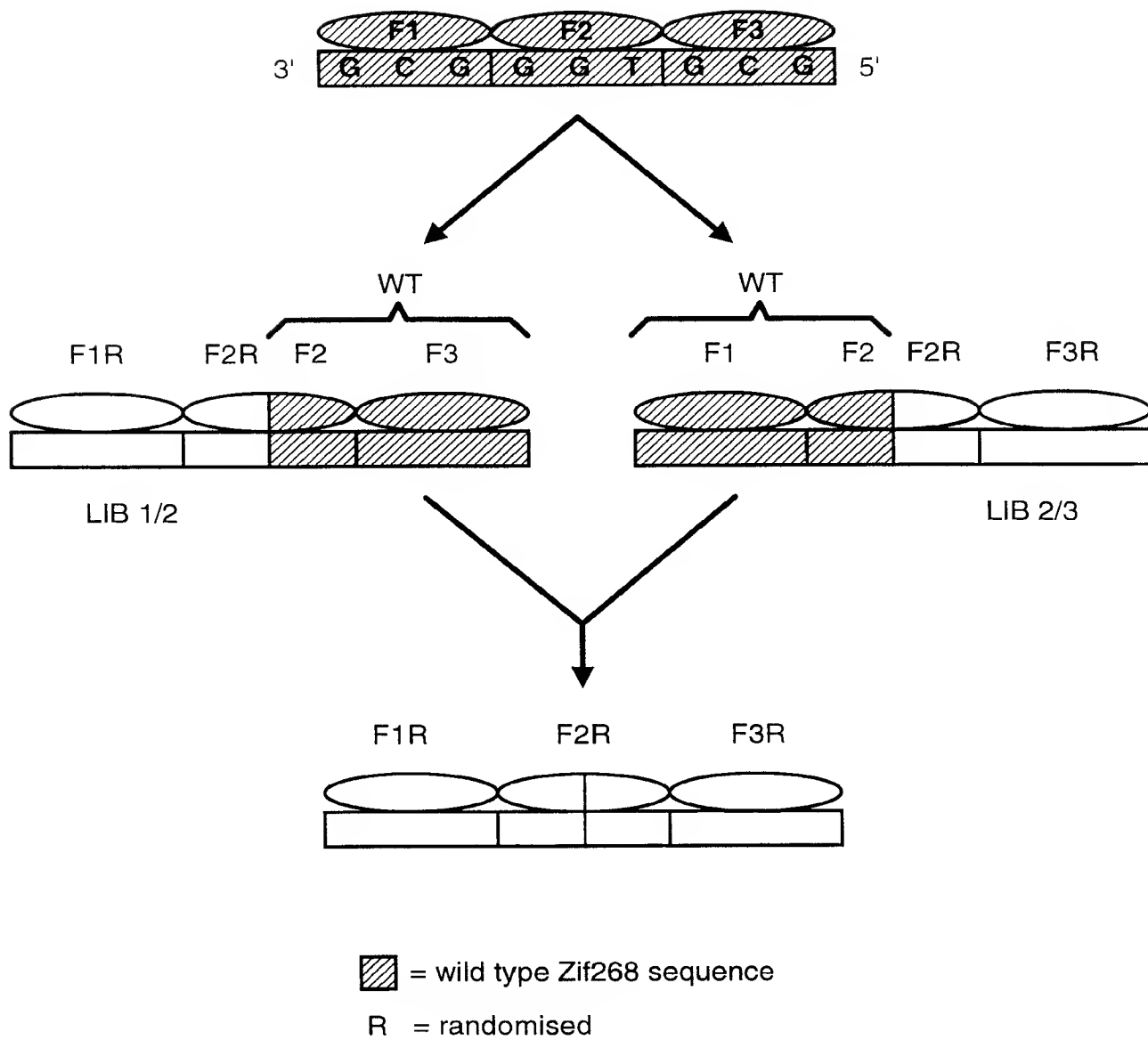


FIG.6

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 98/01510

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/10 C12N15/12 C07K1/04 C12N15/62 C12Q1/68
C07K14/47 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12Q C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	M. ISALAN ET AL: "Synergy between adjacent zinc fingers in sequence-specific DNA recognition" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 94, 27 May 1997, pages 5617-5621, XP002075337 WASHINGTON US see the whole document ---	1-25
X	CHOO Y ET AL: "DESIGNING DNA-BINDING PROTEINS ON THE SURFACE OF FILAMENTOUS PHAGE" CURRENT OPINION IN BIOTECHNOLOGY, August 1995, pages 431-436, XP000652427 cited in the application	1-6
Y	see the whole document see page 432, column 1 ---	10-25
-/--		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search

1 September 1998

Date of mailing of the international search report

30/09/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Cervigni, S

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 98/01510

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>WO 96 06166 A (MEDICAL RES COUNCIL ;CHOO YEN (SG); KLUG AARON (GB); GARCIA ISIDRO) 29 February 1996 cited in the application see the whole document see page 5-6 see page 9, paragraph 3 see page 15, paragraph 2 see figures 2,7,15; tables 1,2 -----</p>	10-25

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 98/01510

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9606166 A	29-02-1996	AU 3229195 A	14-03-1996
		CA 2196419 A	29-02-1996
		EP 0781331 A	02-07-1997
		JP 10504461 T	06-05-1998
<hr/>			

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
7 December 2000 (07.12.2000)

PCT

(10) International Publication Number
WO 00/73434 A1(51) International Patent Classification⁷: C12N 15/10,
15/63, 15/82, C07K 14/47, C12Q 1/68, G01N 33/50,
A01H 5/00[GR/GB]; MRC Laboratory of Molecular Biology, Medi-
cal Research Council Centre, Hills Road, Cambridge CB2
2QH (GB). ULLMAN, Christopher, Graeme [GB/GB];
7 Gateway Mews, Shacklewell Lane, London E8 2DF
(GB).

(21) International Application Number: PCT/GB00/02071

(22) International Filing Date: 30 May 2000 (30.05.2000)

(74) Agents: MASCHIO, Antonio et al.; D Young & Co., 21
New Fetter Lane, London EC4A 1DA (GB).

(25) Filing Language: English

(26) Publication Language: English

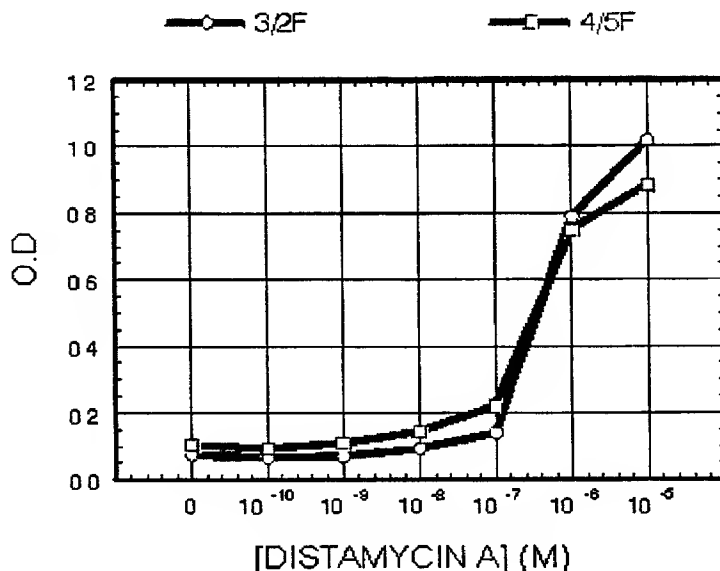
(30) Priority Data:
9912635.1 28 May 1999 (28.05.1999) GB
0001578.4 24 January 2000 (24.01.2000) GB(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE,
DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU,
ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO,
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(71) Applicant (*for all designated States except US*): GENDAQ
LIMITED [GB/GB]; 1-3 Burtonhole Lane, Mill Hill, Lon-
don NW7 1AD (GB).(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): CHOO, Yen

[Continued on next page]

(54) Title: GENE SWITCHES



(57) Abstract: A method is provided of selecting a gene switch, which gene switch comprises (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand, which method comprises: (a) contacting one or more candidate target DNA molecule(s) with one or more candidate DNA binding molecules, in the presence of one or more DNA binding ligands, wherein at least one of the candidate DNA binding molecules comprises a non-naturally occurring DNA binding domain; (b) selecting a complex comprising a candidate target DNA, a DNA binding molecule and a DNA binding ligand; (c) isolating and/or identifying the unknown components of the complex; (d) comparing the binding of the DNA

binding molecule component of the complex to the target DNA component of the complex in the presence and absence of the DNA binding ligand component of the complex; and (e) selecting complexes where said binding differs in the presence and absence of the DNA binding ligand component.

WO 00/73434 A1

**Published:**

- *With international search report.*
- *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Gene Switches

Field of the Invention

This invention relates to molecular gene switches that use molecules capable of binding a specific DNA sequence in a ligand-dependent manner where the ligand itself is capable of binding DNA. Moreover, this invention relates to methods for the identification of said ligand-dependent DNA binding molecules.

Background to the Invention

Gene switches are currently of great interest to those wishing to control timing and/or dosage of gene expression. Various gene switches have been developed in the prior art. Most of these prior art switches are derived from gene regulatory proteins. In these systems, the switching ligand binds to the protein, inducing a protein conformational change that affects DNA binding.

It is often the case that a gene's expression is affected by one or more different protein(s). Diverse proteins may influence expression of the same gene. Said protein(s) may be present in a first cell or cell type, but these protein(s) may be absent from a second cell or cell type. Therefore, a molecule which affects only a single known regulatory protein will not have any effect on the expression of the same gene in a cell where this particular regulatory protein is not expressed, or is otherwise sequestered. Thus, one of the difficulties of the prior art is that a protein-binding switching molecule will have no effect on the expression of a gene if the particular protein to which the switching molecule binds is not present.

Similarly, a gene's expression may be affected by numerous different proteins in different cells or cell types. A molecule which affects only a single known regulatory protein will not have any effect on the expression of the same gene in a cell in which its expression is controlled by a different protein or proteins. Therefore, one of the difficulties in the prior art is that a plurality of switching molecules may be required in order to modulate or switch the expression of a single gene.

Therefore, in order to effect switching of gene expression at a given DNA sequence, independently of the particular activator protein, it is desirable to target the DNA. Further, custom DNA binding proteins would benefit from switches; if these could be designed to
5 interact with DNA, there would be a greater freedom in the design of said proteins.

There are numerous polypeptide modifications which are known to affect their interaction with a broad spectrum of molecules such as nucleic acids, polypeptides (both intra- and inter-molecularly), other macromolecular structures such as membranes, small molecules,
10 ions, or other entities. Clearly, it is a problem that polypeptide modifications may compromise the binding of prior art switching molecules to their polypeptide targets.

The present invention seeks to overcome such difficulties.

15 Aspects of the present invention are set out in the claims and are described below.

Summary of the Invention

In a first aspect, the present invention provides a method of selecting a gene switch, which
20 gene switch comprises (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand, which method comprises:

- (a) contacting one or more candidate target DNA molecule(s) with one or more candidate DNA binding molecules, in the presence of one or more DNA binding ligands,
25 wherein at least one of the candidate DNA binding molecules comprises a non-naturally occurring DNA binding domain;
- (b) selecting a complex comprising a candidate target DNA, a DNA binding molecule and a DNA binding ligand;
- (c) isolating and/or identifying the unknown components of the complex;
- 30 (d) comparing the binding of the DNA binding molecule component of the complex to the target DNA component of the complex in the presence and absence of the DNA binding ligand component of the complex; and

(e) selecting complexes where said binding differs in the presence and absence of the DNA binding ligand component.

5 Preferably the DNA binding molecules are provided as a plurality of DNA binding molecules, more preferably as a library of DNA binding molecules. Where only one DNA binding molecule is included in the screen, the DNA binding molecule comprises a non-naturally occurring DNA binding domain. The term "a non-naturally occurring DNA binding domain" means that the DNA binding domain does not occur in nature, even as part of a larger molecule, and has been obtained by deliberate mutagenesis procedures or *de*
10 *novo* design techniques.

Preferably the target DNA is provided as a plurality of DNA sequences, more preferably as a library of DNA sequences, said sequences being related to one another by sequence homology.

15 In one embodiment, a plurality of candidate DNA binding ligands are used, in which case is preferred to use one target DNA.

Typically one of the components isolated and/or identified in step (c) is a DNA binding
20 ligand component or a DNA binding molecule component.

In a preferred embodiment of the first aspect of the invention, the selected DNA binding molecule component has a higher affinity for the target DNA in the presence of the DNA binding ligand component than in the absence of the DNA binding ligand component.

25 Alternatively, the selected DNA binding molecule component has a higher affinity for the target DNA in the absence of the DNA binding ligand component than in the presence of the DNA binding ligand component.

30 In a highly preferred embodiment, the candidate DNA binding molecules are provided as a phage display library.

The method of the present invention may be used to select a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA binding ligand.

5 The method of the present invention may also be used to select a target DNA to which binds a DNA binding molecule in a manner modulatable by a DNA binding ligand.

The method of the present invention may also further be used to select a DNA binding ligand that modulates binding of a DNA binding molecule to a target DNA.

10 Generally, the DNA binding ligand and the DNA binding molecule are different

In a preferred aspect of the invention, said candidate molecules are polypeptides. In a more preferred embodiment, said candidate molecules are polypeptides at least partly derived from transcription factors. In an even more preferred embodiment, said candidate
15 molecules are derived from zinc finger transcription factors.

Advantageously, the candidate DNA binding molecules are provided as a phage display library.

20 In a preferred aspect of the invention, the DNA binding ligand is selected from Distamycin A, Actinomycin D and echinomycin.

In another aspect, the invention relates a gene switch comprising (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a
25 manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand. In particular, the present invention relates to DNA binding molecules and/or DNA binding ligands and/or target DNA obtainable by the methods disclosed herein.

30 The present invention also provides a method for engineering a novel class of gene switches in which a DNA binding ligand affects or modulates the interaction of a DNA binding molecule (for example phage displayed polypeptide), with its target DNA. In a preferred aspect, the present invention relates to the selection of DNA binding polypeptides which recognise a particular DNA sequence or structure. Preferably, said method may

include selection of phage displayed polypeptides that bind a DNA target in the presence or absence of one or more DNA binding ligands. Of the phage displayed polypeptides which are selected under these conditions, some may bind the DNA with higher affinity in the presence of ligand, whereas others may bind the DNA with higher affinity in the absence of
5 ligand.

The gene switches and components thereof can be used in methods of regulating gene expression. Accordingly, the present invention also provides a method of modulating the expression of one or more genes, said method comprising administering a DNA binding
10 molecule and DNA binding ligand selected according to the method of the invention to a cell wherein the regulatory sequences of said genes comprise a target DNA selected according to the method of the invention.

The present invention also provides a method of modulating the expression of one or more
15 nucleotide sequences of interest in a host cell which host cell comprises a nucleic acid sequence capable of directing the expression of a DNA binding molecule and a target DNA sequence to which the DNA binding molecule binds in a manner modulatable by a DNA binding ligand which method comprises administering said DNA binding ligand to the cell and wherein the DNA binding molecule is heterologous to the host cell.

20 Preferably the host cell is a plant cell. More preferably the plant cell is part of a plant and the target sequence is part of a regulatory sequence to which the nucleotide sequence of interest is operably linked, said regulatory sequence being preferentially active in the male or female organs of the plant.

25 In a further aspect there is provided the use of a DNA binding molecule selected by the method of the invention in a method of regulating transcription from a DNA sequence comprising a target DNA to which the DNA binding molecule binds in a manner modulatable by a DNA binding ligand.

30 Also provided is the use of a DNA binding ligand selected by the method of the invention in a method of regulating transcription from a DNA sequence comprising a target DNA to

which a DNA binding molecule binds in a manner modulatable by the DNA binding ligand.

Also provided is the use of a target DNA selected by the method of the invention in a method of regulating transcription from a DNA sequence comprising the target DNA to which a DNA binding molecule binds in a manner modulatable by a DNA binding ligand.

In another aspect, the present invention provides a non human transgenic organism comprising a target DNA sequence and a nucleic acid sequence capable of directing the expression of a DNA binding molecule which binds to the target DNA in a manner modulatable by a DNA binding ligand wherein the target DNA sequence and/or nucleic acid sequence are heterologous to the organism.

Preferably the transgenic non-human organism is a plant.

Detailed Description of the Invention

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. which are incorporated herein by reference), chemical methods, pharmaceutical formulations and delivery and treatment of patients.

The term 'modulatable by' is used to indicate that binding of the DNA binding molecule to the DNA can be modulated or affected by the DNA binding ligand. In other words, the DNA binding ligand can modulate, affect, regulate, adjust, alter, or vary the binding of the DNA binding molecule to the DNA.

The term 'isolating' in the context of the invention, refers to the act of removing one or more components or molecules from a sample of candidate molecules which are used in the methods disclosed herein.

5

The term 'complex' is used to describe an association between a DNA and one or more molecules as defined herein.

The term "gene switch" is used herein to describe a multiple component system comprising
10 (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand. The DNA binding molecule may or may not comprise a transcriptional effector domain, especially when part of the assay procedure. However, since ultimately the gene switch will be used to regulate transcription from one or more promoters, the DNA binding
15 molecule may need to be modified to include a transcriptional activator or repressor domain, if one is not already present.

The terms "DNA binding molecule", "DNA binding ligand" and "target DNA" are used extensively herein. However other types of nucleic acids other than DNA may be relevant.
20 Consequently, it is intended that in general the above terms can be replaced with the terms "nucleic acid binding molecule", "nucleic acid binding ligand" and "target nucleic acid", respectively. Nucleic acids will in general be RNA or DNA, double stranded or single stranded. RNA is preferably at least partially double-stranded in the context of the present invention. However, in a preferred aspect of the invention, references to "DNA" mean
25 deoxyribonucleic acid in a literal sense.

A. DNA binding molecules

The term 'DNA binding molecule' includes any molecule which is capable of binding or
30 associating with DNA. This binding or association may be via covalent bonding, via ionic bonding, via hydrogen bonding, via Van-der-Waals bonding, or via any other type of reversible or irreversible association.

The term 'molecule' is used herein to refer to any atom, ion, molecule, macromolecule (for example polypeptide), or combination of such entities. The term 'ligand' is used interchangeably with the term 'molecule'. Molecules according to the invention may be free in solution, or may be partially or fully immobilised. They may be present as discrete entities, or may be complexed with other molecules. Preferably, molecules according to the invention include polypeptides displayed on the surface of bacteriophage particles. More preferably, molecules according to the invention include libraries of polypeptides presented as integral parts of the envelope proteins on the outer surface of bacteriophage particles. Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids, and not for stop codons.

The term 'candidate DNA binding molecules' is used to describe any one or more molecule(s) as defined above which may or may not be capable of binding DNA. The capability of said molecules to bind DNA may or may not be modulatable by a DNA binding ligand. The latter of these properties may be investigated by the methods of this invention. Preferably, candidate DNA binding molecules comprise a plurality of, or a library of polypeptides. More preferably, these polypeptides are, or are derived from, DNA binding proteins such as DNA repair enzymes, polymerases, recombinases, methylases, restriction enzymes, replication factors, histones, or DNA binding structural proteins such as chromosomal scaffold proteins; even more preferably said polypeptides are derived from transcription factors. 'Derived from' means that the candidate DNA binding molecules preferably comprise one or more of; transcription factors, fragment(s) of transcription factors, sequences homologous to transcription factors, or polypeptides which have been fully or partially randomised from a starting sequence which is a transcription factor, a fragment of a transcription factor, or homologous to a transcription factor. Most preferably, candidate DNA binding molecules comprise polypeptides which are at least 40% homologous, more preferably at least 60% homologous, even more preferably at least 75% homologous or even more, for example 85 %, or 90 %, or even more than 95% homologous to one or more transcription factors, using one of the homology calculation algorithms defined below.

Candidate DNA binding molecules may comprise, among other things, DNA binding part(s) of any protein(s), for example zinc finger transcription factors, Zif268, ATF family transcription factors, ATF1, ATF2, bZIP proteins, CHOP, NF- κ B, TATA binding protein (TBP), MDM, c-jun, elk, serum response factor (SRF), ternary complex factor (TCF); KRÜPPEL, Odd Skipped, even skipped and other *D.melanogaster* transcription factors; yeast transcription factors such as GCN4, the GAL family of galactose-inducible transcription factors; bacterial transcription factors or repressors such as *lacI*^q, or fragments or derivatives thereof. Derivatives would be considered by a person skilled in the art to be functionally and/or structurally related to the molecule(s) from which they are derived, for example through sequence homology of at least 40%.

The candidate DNA binding molecules may be non-randomised polypeptides, for example 'wild-type' or allelic variants of naturally occurring polypeptides, or may be specific mutant(s), or may be wholly or partially randomised polypeptides, preferably structurally related to DNA binding proteins as described herein.

In a highly preferred embodiment, these polypeptide candidate DNA binding molecules are displayed on the surface of bacteriophage particles, and are preferably partially randomised zinc-finger type transcription factors, preferably retaining at least 40% homology (as described herein) to zinc-finger type transcription factors.

In some cases, sequence homology may be considered in relation to structurally important residues, or those residues which are known or suspected of being evolutionarily conserved. In such instances, residues known to be variable or non-essential for a particular structural conformation may be discounted from the homology calculation. For example, as explained herein, zinc fingers are known to have certain residues which are important for the formation of the three-dimensional zinc finger structure. In these cases, homology may be considered over about seven of said important amino acid residues amongst approximately thirty residues which may comprise the whole finger structure.

As used herein, the term homology may refer to structural homology. Structural homology may be estimated by comparing the structural RMS deviation of the main part of the carbon atom backbone of two or more molecules. Preferably, the molecules may be considered

structurally homologous if the deviation is 5Å or less, preferably 3Å or less, more preferably 1.5Å or less. Structurally homologous molecules will not necessarily show significant sequence homology.

- 5 Candidate DNA binding molecules, as defined above, may be prescreened prior to being tested in the methods of the invention using routine assays known in art for determining the binding of molecules to nucleic acids so as to eliminate molecules that do not bind DNA. For example, a candidate DNA binding molecule, preferably a library of candidate DNA binding molecules, are contacted with nucleic acid and binding determined. The nucleic
- 10 acids may for example be labelled with a detectable label, such as a fluorophore/flurochrome, such that after a wash step binding can be determined easily, for example by monitoring fluorescence. Other methods for measuring binding to DNA are set out in section E. Below.
- 15 The nucleic acid with which the candidate binding ligands are contacted may be non-specific nucleic acids, such as a random oligonucleotide library or sonicated genomic DNA and the like. Alternatively, a specific sequence may be used or partially randomised library of sequences.
- 20 Preferably, the DNA binding molecules of the invention may bind the target nucleic acid with different affinity in the presence or in the absence of ligand. The binding to the nucleic acid may be enhanced by the presence of the ligand (i.e. bind with a higher affinity in the presence of ligand), or may be reduced in the presence of ligand (i.e. bind with a lower affinity in the presence of ligand). In the case where association of the DNA binding
- 25 molecule(s) with the target nucleic acid is enhanced by the presence of ligand, said association may be additive with the binding of the ligand, or may be synergistic with the binding of the ligand, or may affect the binding in another way. If the binding is synergistic with the binding of the ligand, said binding may be either wholly or partly dependent on the presence of the ligand. Preferably, the characteristics of binding may be
- 30 such that the DNA binding molecule(s) may be eluted by addition of an excess of the DNA binding ligand.

DNA binding molecules according to the invention are preferably polypeptide sequences, optionally encoded by nucleic acid sequences. Fragments, mutants, alleles and other derivatives of the molecules of the invention preferably retain substantial homology with said sequence(s). As used herein, "homology" means that the two entities share sufficient characteristics for the skilled person to determine that they are similar. Preferably, homology is used to refer to sequence identity. Thus, the derivatives of said DNA binding molecules of the invention preferably retain substantial sequence identity with said molecules.

In the context of the present invention, a homologous sequence is taken to include any sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical over at least 5, preferably 8, 10, 15, 20, 30, 40 or even more residues or bases with the molecules (ie. the sequences thereof) of the invention, for example as shown in the sequence listing herein. In particular, homology should typically be considered with respect to those regions of the molecule(s) which may be known to be functionally important rather than non-essential neighbouring sequences. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues (for example less than 50 contiguous amino acids).

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a

large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting “gaps” in the sequence alignment to try to
5 maximise local homology.

However, these more complex methods assign “gap penalties” to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences -
10 will achieve a higher score than one with many gaps. “Affine gap costs” are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default
15 values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

Calculation of maximum % homology therefore firstly requires the production of an optimal
20 alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.*, 1984, Nucleic Acids Research 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al.*, 1999 *ibid* – Chapter 18), FASTA (Atschul *et al.*, 1990, J. Mol.
25 Biol., 403-410) and the GENWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

Although the final % homology can be measured in terms of identity, the alignment process
30 itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of

programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

5

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

- 10 DNA binding molecules according to the invention may include any atom, ion, molecule, macromolecule (for example polypeptide), or combination of such entities that are capable of binding to nucleic acids, such as DNA. Advantageously, molecules according to the invention may include families of polypeptides with known or suspected nucleic acid binding motifs. These may include for example zinc finger proteins (see below).
- 15 Molecules according to the invention may also include helix-turn-helix proteins, homeodomains, leucine zipper proteins, helix-loop-helix proteins or β -sheet motifs which are well known to a person skilled in the art.

- According to the invention, DNA binding motifs of one or more known or suspected
- 20 nucleic acid binding polypeptide(s) may advantageously be randomised, in order to provide libraries of candidate nucleic acid binding molecules.

- Crystal structures may advantageously be used in selecting or predicting the relevant DNA binding regions of nucleic acid binding proteins by methods known in the art.

25

DNA binding regions of proteins within the same structural family are often conserved or homologous to one another, for example zinc finger α -helices, the leucine zipper basic region, homeodomain helix 3.

- 30 General considerations and rules governing the binding of several polypeptide families to nucleic acids are set out in the literature, e.g. in (Suzuki *et al.*, 1994:PNAS vol 91 pp

12357-61). Nucleic acid binding criteria for zinc fingers as preferred DNA binding molecules according to the present invention are set out in this application (see above).

It is also envisaged that the methods of the present invention could be advantageously applied to the selection of ligand-modulatable DNA binding molecules from other families of transcription factors, for example from the helix-turn-helix (HTH) family and/or from the probe helix (PH) family, and/or from the C4 Zinc-binding family (which includes the hormone receptor (HR) family), from the Gal4 family, from the c-myb family, from other zinc finger families, or from any other family of DNA binding proteins known to one skilled in the art.

One or more polypeptides from one or more of these families could be advantageously randomised to provide a library of candidate molecules for use in the methods of the invention. Preferably, the amino acid residues known to be important for nucleic acid binding could be randomised. However, it may be desirable to randomise other regions of the DNA binding molecule since alterations to the amino acid sequence outside of those elements of secondary structure that present amino acids that contact the DNA are likely to cause conformational changes that may affect the DNA binding properties of the molecule.

For example, randomisation may involve alteration of zinc finger polypeptides, said alteration being accomplished at the DNA or protein level. Mutagenesis and screening of zinc finger polypeptides may be achieved by any suitable means. Preferably, the mutagenesis is performed at the nucleic acid level, for example by synthesising novel genes encoding mutant polypeptides and expressing these to obtain a variety of different proteins. Alternatively, existing genes can themselves be mutated, such as by site-directed or random mutagenesis, in order to obtain the desired mutant genes.

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the protein of interest. A number of methods for site-directed mutagenesis are known in the art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see "PCR Protocols: A guide to methods and applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York, 1990). Preferably, the

commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the manufacturer's instructions.

Randomisation of the zinc finger binding motifs is preferably directed to those amino acid residues where the code provided herein gives a choice of residues (see below). For example, positions +1, +5 and +8 are advantageously randomised, whilst preferably avoiding hydrophobic amino acids; positions involved in binding to the nucleic acid, notably -1, +2, +3 and +6, may be randomised also, preferably within the choices provided by the rules of the present invention.

Screening of the proteins produced by mutant genes is preferably performed by expressing the genes and assaying the binding ability of the protein product. A simple and advantageously rapid method by which this may be accomplished is by phage display, in which the mutant polypeptides are expressed as fusion proteins with the coat proteins of filamentous bacteriophage, such as the minor coat protein pII of bacteriophage m13 or gene III of bacteriophage Fd, and displayed on the capsid of bacteriophage transformed with the mutant genes. The target nucleic acid sequence is used as a probe to bind directly to the protein on the phage surface and select the phage possessing advantageous mutants, by affinity purification. The phage are then amplified by passage through a bacterial host, and subjected to further rounds of selection and amplification in order to enrich the mutant pool for the desired phage and eventually isolate the preferred clone(s). Detailed methodology for phage display is known in the art and set forth, for example, in US Patent 5,223,409; Choo and Klug, (1995) *Current Opinions in Biotechnology* 6:431-436; Smith, (1985) *Science* 228:1315-1317; and McCafferty *et al.*, (1990) *Nature* 348:552-554; all incorporated herein by reference. Vector systems and kits for phage display are available commercially, for example from Pharmacia.

Specific peptide ligands such as zinc finger polypeptides may moreover be selected for binding to targets by affinity selection using large libraries of peptides linked to the C-terminus of the lac repressor LacI (Cull *et al.*, (1992) *Proc Natl Acad Sci U S A*, 89, 1865-9). When expressed in *E. coli* the repressor protein physically links the ligand to the encoding plasmid by binding to a lac operator sequence on the plasmid.

An entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, (1994) Proc Natl Acad Sci U S A, 91, 9022-6) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them. Furthermore, polypeptides may be partitioned in physical compartments for example wells of an *in vitro* dish, or subcellular compartments, or in small fluid particles or droplets such as emulsions; further teachings on this topic may be found in Griffith *et al.*, (see WO 99/02671).

A library for use in the invention may be randomised at those positions for which choices are given in the rules of the first embodiment of the present invention. The rules set forth above allow the person of ordinary skill in the art to make informed choices concerning the desired codon usage at the given positions.

The recognition helix of PH family polypeptides contains conserved Arg/Lys residues which are important structural elements involved in the binding of phosphates in the nucleic acid. Base specificity is attributed to amino acids 1, 4, 5 and 8 of the helix. These residues could be advantageously varied, for example amino acid 1 could be selected from Asn, Asp, His, Val, Ile to provide the possibility of binding to A, C, G, or T. Similarly, amino acid 4 could be selected from Asn, Asp, His, Val, Ile, Gln, Glu, Arg, Lys, Met, or Leu to provide the possibility of binding to A,C,G or T. Preferably, the rules laid out in (Suzuki *et al.*, 1994: PNAS vol 91 pp 12357-61) would be used in order to randomise those amino acids which affect interaction of the molecule with the nucleic acid, whether in a base specific manner, or via binding to the phosphate backbone, thereby producing a library of candidate nucleic acid binding molecules for use in the methods of the invention.

Similarly, polypeptide molecules of the helix-turn-helix family could be randomised to produce a library of candidate molecules, at least some of which may preferably be capable of binding nucleic acid in a ligand-dependent manner when used in the methods of the present invention. In particular, amino acids 1, 2, 5 and 6 are known to be conserved and function in base-specific nucleic acid binding in HTH motifs. Therefore, at least amino acids 1, 2, 5 or 6 would preferably be randomised so as to produce molecules for use according to the present invention. More preferably, amino acids 1, 5 and 6 could be selected from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys or Leu, and amino acid 2

could be selected from from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys, Leu, Cys, Ser, Thr, or Ala.

Another family of transcription factors which may be advantageously employed in the methods of the current invention are the C4 family which includes hormone receptor type transcription factors. It is envisaged that polypeptides of this family could advantageously be used to provide candidate molecules for use in selecting nucleic acid binding molecules whose association with nucleic acid is modulatable by a nucleic acid binding ligand. Amino acids 1, 4, 5 and 9 of the C4 motif are known to be involved in contacting the DNA, and therefore these residues would preferably be altered to provide a plurality of different molecules which may bind DNA in a ligand dependent manner. Preferably, amino acids 1 and 5 could be selected from from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys or Leu, and amino acids 4 and 9 could be selected from Gln, Glu, Arg, Lys, Leu or Met.

Particularly preferred examples of DNA binding molecules are Cys2-His2 zinc finger binding proteins which, as is well known in the art, bind to target nucleic acid sequences via α -helical zinc metal atom co-ordinated binding motifs known as zinc fingers. Each zinc finger in a zinc finger nucleic acid binding protein is responsible for determining binding to a nucleic acid triplet, or an overlapping quadruplet, in a nucleic acid binding sequence. Preferably, there are 2 or more zinc fingers, for example 2, 3, 4, 5 or 6 zinc fingers, in each binding protein. Advantageously, there are 3 zinc fingers in each zinc finger binding protein.

Thus, in one embodiment, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence, wherein binding is via a zinc finger DNA binding motif of the polypeptide, and wherein said binding is modulatable by a DNA binding ligand.

All of the DNA binding residue positions of zinc fingers, as referred to herein, are numbered from the first residue in the α -helix of the finger, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α -helix in a Cys2-His2 zinc finger polypeptide. Residues referred to as "++" are residues present in an

adjacent (C-terminal) finger. Where there is no C-terminal adjacent finger, “++” interactions do not operate.

The present invention is in one aspect concerned with the production of what are essentially artificial DNA binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term “amino acid”, particularly in the context where “any amino acid” is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues or mimetics of the defined amino acids.

The α -helix of a zinc finger binding protein aligns antiparallel to the nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N terminal to C-terminal sequence of the zinc finger. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus to C-terminus, the result is that when a nucleic acid sequence and a zinc finger protein are aligned according to convention, the primary interaction of the zinc finger is with the - strand of the nucleic acid. since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain fingers, such as finger 4 of the protein GLI, bind to the + strand of nucleic acid: see Suzuki *et al.*, (1994) NAR 22:3397-3405 and Pavletich and Pabo, (1993) Science 261:1701-1707. The incorporation of such fingers into DNA binding molecules according to the invention is envisaged.

The present invention may be integrated with the rules set forth for zinc finger polypeptide design in our copending European or PCT patent applications having publication numbers; WO 98/53057, WO 98/53060, WO 98/53058, WO 98/53059, describe improved techniques for designing zinc finger polypeptides capable of binding desired nucleic acid sequences. In combination with selection procedures, such as phage display, set forth for

example in WO 96/06166, these techniques enable the production of zinc finger polypeptides capable of recognising practically any desired sequence.

In a preferred aspect, therefore, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, and wherein binding to each base of the triplet by an α -helical zinc finger DNA binding motif in the polypeptide is determined as follows:

- a) if the 5' base in the triplet is G, then position +6 in the α -helix is Arg and/or position ++2 is Asp;
- b) if the 5' base in the triplet is A, then position +6 in the α -helix is Gln or Glu and ++2 is not Asp;
- c) if the 5' base in the triplet is T, then position +6 in the α -helix is Ser or Thr and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;
- d) if the 5' base in the triplet is C, then position +6 in the α -helix may be any amino acid, provided that position ++2 in the α -helix is not Asp;
- e) if the central base in the triplet is G, then position +3 in the α -helix is His;
- f) if the central base in the triplet is A, then position +3 in the α -helix is Asn;
- g) if the central base in the triplet is T, then position +3 in the α -helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;
- h) if the central base in the triplet is 5-meC, then position +3 in the α -helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;
- i) if the 3' base in the triplet is G, then position -1 in the α -helix is Arg;
- j) if the 3' base in the triplet is A, then position -1 in the α -helix is Gln and position +2 is Ala;
- k) if the 3' base in the triplet is T, then position -1 in the α -helix is Asn; or position -1 is Gln and position +2 is Ser;
- l) if the 3' base in the triplet is C, then position -1 in the α -helix is Asp and Position +1 is Arg; where the central residue of a target triplet is C, the use of Asp at position +3 of a zinc finger polypeptide allows preferential binding to C over 5-meC.

The foregoing represents a set of rules which permits the design of a zinc finger binding protein specific for any given target DNA sequence.

- 5 A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA) 85:99-102; Lee *et al.*, (1989) Science 245:635-637; see International patent applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

10

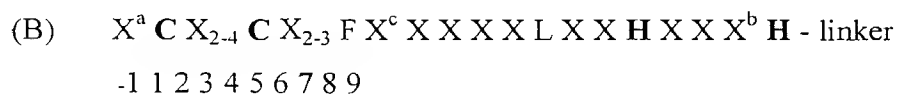
In general, a preferred zinc finger framework has the structure:



- 15 where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:

20



- 25 wherein X (including X^a , X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α -helix.

- 30 Modifications to this representation may occur or be effected without necessarily abolishing zinc finger function, by insertion, mutation or deletion of amino acids. For example it is known that the second His residue may be replaced by Cys (Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances

be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for the zinc finger are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α -helix co-ordinated by a zinc atom which contacts four Cys or His residues, does not alter. As used herein, structures (A) and (B) above are taken as an exemplary structure representing all zinc finger structures of the Cys2-His2 type.

Preferably, X^a is $^F/Y-X$ or $P-^F/Y-X$. In this context, X is any amino acid. Preferably, in this context X is E, K, T or S. Less preferred but also envisaged are Q, V, A and P. The remaining amino acids remain possible.

Preferably, X_{2-4} consists of two amino acids rather than four. The first of these amino acids may be any amino acid, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these amino acids is preferably E, although any amino acid may be used.

Preferably, X^b is T or I. Preferably, X^c is S or T.

Preferably, X_{2-3} is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

Preferably, the linker is T-G-E-K or T-G-E-K-P.

As set out above, the major binding interactions occur with amino acids -1, +3 and +6. Amino acids +4 and +7 are largely invariant. The remaining amino acids may be essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr. Preferably, position ++2 is any amino acid, and preferably serine, save where its nature is dictated by its role as a ++2 amino acid for an N-terminal zinc finger in the same nucleic acid binding molecule.

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger DNA binding motif which will bind specifically to a given target DNA triplet.

5 The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins
10 which are capable of binding to every defined target DNA triplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the
15 first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target triplet.

In a further aspect of the present invention, there is provided a method for preparing a DNA
20 binding protein of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence in a manner modulatable by a DNA binding ligand, comprising the steps of:

- a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and
25
- b) mutating at least one of positions -1, +3, +6 (and ++2) of the finger as required by a method according to the present invention.

In general, naturally occurring zinc fingers may be selected from those fingers for which
30 the DNA binding specificity is known. For example, these may be the fingers for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack

(Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

Consensus zinc finger structures may be prepared by comparing the sequences of known zinc fingers, irrespective of whether their binding domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

The consensus structures are derived from the consensus provided by Krizek *et al.*, (1991) J. Am. Chem. Soc. 113: 4518-4523 and from Jacobs, (1993) PhD thesis, University of Cambridge, UK. In both cases, the linker sequences described above for joining two zinc finger motifs together, namely TGEK or TGEKP can be formed on the ends of the consensus. Thus, a P may be removed where necessary, or, in the case of the consensus terminating T G, E K (P) can be added.

When the nucleic acid specificity of the model finger selected is known, the mutation of the finger in order to modify its specificity to bind to the target DNA may be directed to residues known to affect binding to bases at which the natural and desired targets differ. Otherwise, mutation of the model fingers should be concentrated upon residues -1, +3, +6 and ++2 as provided for in the foregoing rules.

In order to produce a binding protein having improved binding, moreover, the rules provided by the present invention may be supplemented by physical or virtual modelling of the protein/DNA interface in order to assist in residue selection.

In a second embodiment, the invention provides a method for producing a zinc finger polypeptide capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, comprising:

- a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α -helix of the zinc finger polypeptides;
- 5 b) displaying the library in a selection system and screening it against a target DNA sequence;
- c) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence in the presence/absence of DNA binding ligand;
- 10 d) selecting those members of the library isolated in (c) which bind the target nucleic acid sequence with different affinities in the presence and absence of the DNA binding ligand.

Methods for the production of libraries encoding randomised polypeptides are known in the
15 art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids as set forth in the rules above.

Zinc finger polypeptides may be designed which specifically bind to nucleic acids
20 incorporating the base U, in preference to the equivalent base T.

In a further preferred aspect, the invention comprises a method for producing a zinc finger polypeptide capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, comprising:

- 25 a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides each possessing more than one zinc fingers, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α -helix in a first zinc finger and at one or more of the positions encoding
30 residues -1, 2, 3 and 6 of the α -helix in a further zinc finger of the zinc finger polypeptides;
- b) displaying the library in a selection system and screening it against a target DNA sequence;

- c) assessing the affinity of the DNA binding molecules for the target DNA in the presence and absence of the DNA binding ligand, and
- d) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence with different affinities in the presence and absence of DNA binding ligand.

In this aspect, the invention encompasses library technology described in our copending International patent application WO 98/53057, incorporated herein by reference in its entirety. WO 98/53057 describes the production of zinc finger polypeptide libraries in which each individual zinc finger polypeptide comprises more than one, for example two or three, zinc fingers; and wherein within each polypeptide partial randomisation occurs in at least two zinc fingers.

This allows for the selection of the "overlap" specificity, wherein, within each triplet, the choice of residue for binding to the third nucleotide (read 3' to 5' on the + strand) is influenced by the residue present at position +2 on the subsequent zinc finger, which displays cross-strand specificity in binding. The selection of zinc finger polypeptides incorporating cross-strand specificity of adjacent zinc fingers enables the selection of nucleic acid binding proteins more quickly, and/or with a higher degree of specificity than is otherwise possible.

Zinc finger binding motifs designed according to the invention may be combined into nucleic acid binding polypeptide molecules having a multiplicity of zinc fingers. Preferably, the proteins have at least two zinc fingers. In nature, zinc finger binding proteins commonly have at least three zinc fingers, although two-zinc finger proteins such as Tramtrack are known. The presence of at least three zinc fingers is preferred. Nucleic acid binding proteins may be constructed by joining the required fingers end to end, N-terminus to C-terminus. Preferably, this is effected by joining together the relevant nucleic acid sequences which encode the zinc fingers to produce a composite nucleic acid coding sequence encoding the entire binding protein. The invention therefore provides a method for producing a DNA binding protein as defined above, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

- a) preparing a nucleic acid coding sequence encoding two or more zinc finger binding motifs as defined above, placed N-terminus to C-terminus;
- b) inserting the nucleic acid sequence into a suitable expression vector; and
- 5 c) expressing the nucleic acid sequence in a host organism in order to obtain the DNA binding protein.

A "leader" peptide may be added to the N-terminal finger. Preferably, the leader peptide is MAEEKP.

10

B. Nucleic acid vectors encoding DNA binding proteins

A nucleic acid encoding the DNA binding protein according to the invention can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the person of ordinary skill in the art. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether it is to be used for DNA amplification or for nucleic acid expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

25

30

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV40, polyoma,

adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

- 5 Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast, mammalian or plant cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into
- 10 the host genome. However, the recovery of genomic DNA encoding the DNA binding protein is more complex than that of episomally replicated vector because restriction enzyme digestion is required to excise DNA binding protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.
- 15 Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and
- 20 other toxins. e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

- Selectable markers which may be used in fungal cells, for example yeast cells, include wild-type genes which complement auxotrophic defects in for example the Uracil (eg.
- 25 URA3 gene), Lysine (eg. LYS2 gene), Adenine (eg. ADE2 gene), Methionine (eg. MET3 gene), Histidine (eg. HIS3 gene), Tryptophan (eg. TRP1 gene), Leucine (eg. LEU2 gene) or other metabolic pathways. In addition, counter-selection methods are well known in the art. These enable genes to be selected against by the action of a chemical precursor which is harmless unless converted to a toxic product by the action of one or more gene(s).
- 30 Examples of these include; 5-fluoro-orotic acid, which is converted to a toxic compound by the action of the URA3 gene product; α -amino-adipic acid, which is converted to a toxic compound by the LYS2 gene product; allyl alcohol, which is converted to a toxic compound by alcohol dehydrogenase activity as encoded by the ADH genes, or any other

suitable selective regime known to those skilled in the art. Other selective markers are based on the expression of a gene in a fungus such as yeast which overcomes the metabolic arrest induced by, or toxicity of, a chemical entity which may be added to the growth medium or otherwise presented to the cells. Examples of these may include the KAN gene(s) which confer resistance to antibiotics such as G-418, the HIS3 gene which confers resistance to 3-amino-triazole, or the ADH2 gene which can confer resistance to heavy metal ions such as cadmium, or any other suitable genes which confer resistance to toxic or growth arresting regimes.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript© vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up DNA binding protein nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes the DNA binding protein. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to nucleic acid encoding DNA binding protein. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA

encoding the DNA binding protein by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native DNA binding protein promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of DNA binding protein encoding
5 DNA.

Promoters suitable for use with prokaryotic hosts include, for example, the β -lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been
10 published, thereby enabling the skilled worker operably to ligate them to DNA encoding DNA binding protein, using linkers or adapters to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding the DNA binding protein.

15 Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phage λ or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction
20 with pET vectors, the T7 RNA polymerase is produced from the β -lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible lac UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison,
25 USA). Other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL), vectors containing the trc promoters such as pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) or vectors containing the tac promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (New England Biolabs, MA, USA).

30 Moreover, the DNA binding protein gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body.

The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a *Saccharomyces cerevisiae* gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the α - or α -factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 3-phosphoglycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PH05 promoter is, e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

DNA binding protein gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with DNA binding protein sequence, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding DNA binding protein by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from

mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to DNA binding protein DNA, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector encoding a DNA binding protein according to the invention may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the DNA binding protein gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, or in transgenic animals.

Eukaryotic vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding DNA binding protein.

An expression vector includes any vector capable of expressing DNA binding protein nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding DNA binding protein may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, *et al.*, (1989) NAR 17, 6418).

In a preferred embodiment, the DNA binding protein constructs of the invention are expressed in plant cells under the control of transcriptional regulatory sequences that are

known to function in plants. The regulatory sequences selected will depend on the required temporal and spatial expression pattern of the DNA binding protein in the host plant. Many plant promoters have been characterized and would be suitable for use in conjunction with the invention. By way of illustration, some examples are provided below:

5

A large number of promoters are known in the art which direct expression in specific tissues and organs (*e.g.* roots, leaves, flowers) or in cell types (*e.g.* leaf epidermal cells, leaf mesophyll cells, root cortex cells). For example, the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula *Plant Mol. Bio.* 12: 579-589 (1989)) is green tissue-specific; the *trpA* gene promoter is pith cell-specific (WO 93/07278 to Ciba-Geigy); the TA29 promoter is pollen-specific (Mariani *et al.* *Nature* 347: 737-741 (1990); Mariani *et al.* *Nature* 357: 384-387 (1992)).

Other promoters direct transcription under conditions of presence of light or absence or light or in a circadian manner. For example, the GS2 promoter described by Edwards and Coruzzi, *Plant Cell* 1: 241-248 (1989) is induced by light, whereas the AS1 promoter described by Tsai and Coruzzi, *EMBO J* 9: 323-332 (1990) is expressed only in conditions of darkness.

Other promoters are wound-inducible and typically direct transcription not just on wound induction, but also at the sites of pathogen infection. Examples are described by Xu *et al.* (*Plant Mol. Biol.* 22: 573-588 (1993)); Logemann *et al.* (*Plant Cell* 1: 151-158 (1989)); and Firek *et al.* (*Plant Mol Biol* 22: 129-142 (1993)).

A number of constitutive promoters can be used in plants. These include the Cauliflower Mosaic Virus 35S promoter (US 5,352,605 and US 5,322,938, both to Monsanto) including minimal promoters (such as the -90 or -46 CaMV 35S promoter) linked to other regulatory sequences, the rice actin promoter (McElroy *et al.* *Mol. Gen. Genet.* 231: 150-160 (1991)), and the maize and sunflower ubiquitin promoters (Christensen *et al.* *Plant Mol Biol.* 12: 619-632 (1989); Binet *et al.* *Plant Science* 79: 87-94 (1991)).

Using promoters that direct transcription in the plant species of interest, the DNA binding protein of the invention can be expressed in the required cell or tissue types. For example,

if it is the intention to utilize the DNA binding protein to regulate a gene in a specific cell or tissue type, then the appropriate promoter can be used to direct expression of the DNA binding protein construct.

5 An appropriate terminator of transcription is fused downstream of the selected DNA binding protein containing transgene and any of a number of available terminators can be used in conjunction with the invention. Examples of transcriptional terminator sequences that are known to function in plants include the *nopaline synthase* terminator found in the pBI vectors (Clontech catalog 1993/1994), the E9 terminator from the *rbcS* gene (*ref*), and
10 the *tml* terminator from Cauliflower Mosaic Virus.

A number of sequences found within the transcriptional unit are known to enhance gene expression and these can be used within the context of the current invention. Such sequences include intron sequences which, particularly in monocotyledonous cells, are
15 known to enhance expression. Both intron 1 of the maize *Adhl* gene and the intron from the maize *bronze1* gene have been found to be effective in enhancing expression in maize cells (Callis *et al.* Genes Develop. 1: 1183-1200 (1987)) and intron sequences are frequently incorporated into plant transformation vectors, typically within the non-translated leader.

20 A number of virus-derived non-translated leader sequences have been found to enhance expression, especially in dicotyledonous cells. Examples include the "Ω" leader sequence of Tobacco Mosaic Virus, and similar leader sequences of Maize Chlorotic Mottle Virus and Alfalfa Mosaic Virus (Gallie *et al.* Nucl. Acids Res. 15: 8693-8711 (1987); Shuzeski *et al.* Plant Mol Biol, 15: 65-79 (1990)).
25

The DNA binding proteins of the current invention are targeted to the cell nucleus so that they are able to interact with host cell DNA and bind to the appropriate DNA target in the nucleus and regulate transcription. To effect this, a Nuclear Localization Sequence (NLS)
30 is incorporated in frame with the expressible zinc finger construct. The NLS can be fused either 5' or 3' to the zinc finger encoding sequence.

The NLS of the wild-type Simian Virus 40 Large T-Antigen (Kalderon *et al.* Cell 37: 801-813 (1984); Markland *et al.* Mol. Cell Biol. 7: 4255-4265 (1987)) is an appropriate NLS and has previously been shown to provide an effective nuclear localization mechanism in plants (van der Krol *et al.* Plant Cell 3: 667-675 (1991)). However, several alternative
5 NLSs are known in the art and can be used instead of the SV40 NLS sequence. These include the Nuclear Localization Signals of TGA-1A and TGA-1B (van der Krol *et al.*; Plant Cell 3: 667-675 (1991)).

A variety of transformation vectors are available for plant transformation and the DNA
10 binding protein encoding genes of the invention can be used in conjunction with any such vectors. The selection of vector will depend on the preferred transformation technique and the plant species which is to be transformed. For certain target species, different selectable markers may be preferred.

15 For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable. A number of vectors are available including pBIN19 (Bevan, Nucl. Acids Res. 12: 8711-8721 (1984), the pBI series of vectors, and pCIB10 and derivatives thereof (Rothstein *et al.* Gene 53: 153-161 (1987); WO 95/33818 to Ciba-Geigy).

20 Binary vector constructs prepared for *Agrobacterium* transformation are introduced into an appropriate strain of *Agrobacterium tumefaciens* (for example, LBA 4044 or GV 3101) either by triparental mating (Bevan; Nucl. Acids Res. 12: 8711-8721 (1984)) or direct transformation (Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988)).

25 For transformation which is not *Agrobacterium*-mediated (*i.e.* direct gene transfer), any vector is suitable and linear DNA containing only the construct of interest may be preferred. Direct gene transfer can be undertaken using a single DNA species or multiple DNA species (co-transformation; Schroder *et al.* Biotechnology 4: 1093-1096 (1986)).

30 Particularly useful for practising several embodiments of the present invention are expression vectors that provide for the transient expression of DNA encoding a DNA binding protein in plant cells or mammalian cells. Transient expression usually involves

the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of DNA binding protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying DNA binding protein mutants, to identify
5 potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct
10 sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing DNA binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern
15 blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or *in situ* hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

20 In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryote, yeast and higher eukaryote cells may be used for replicating DNA and producing the DNA binding protein. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E.coli*, e.g. *E.coli* K-12 strains, DH5 α and HB101, or Bacilli. Further
25 hosts suitable for the DNA binding protein encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. *Saccharomyces cerevisiae*. Higher eukaryotic cells include plant cells and animal cells such as insect and vertebrate cells, particularly mammalian cells including human cells, or nucleated cells from other multicellular organisms. In recent years propagation of vertebrate cells in culture (tissue culture) has
30 become a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a multicellular host organism.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of the DNA binding protein-encoding nucleic acid to form the DNA binding protein. The precise amounts of DNA encoding the DNA binding protein may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-mentioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press).

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions whereby the DNA binding protein encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

Transformation of plant cells is normally undertaken with a selectable marker which may provide resistance to an antibiotic or to a herbicide. Selectable markers that are routinely used in transformation include the *nptII* gene which confers resistance to kanamycin (Messing & Vierra *Gene* 19: 259-268 (1982); Bevan *et al.* *Nature* 304: 184-187 (1983)), the *bar* gene which confers resistance to the herbicide phosphinothricin (White *et al.* *Nucl. Acids Res.* 18: 1062 (1990); Spencer *et al.* *Theor. Appl. Genet.* 79: 625-631 (1990)), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochliger & Diggelmann *Mol. Cell Biol.* 4: 2929-2931 (1984)), and the *dhfr* gene which confers resistance to methotrexate (Bourouis *et al.* *EMBO J* 2: 1099-1104 (1983)). More recently, a number of selection systems have been developed which do not rely of selection for resistance to antibiotic or herbicide. These include the inducible isopentyl transferase system described by Kunkel *et al.* (*Nature Biotechnology* 17: 916-919 (1999)).

Although specific protocols may vary from species to species, transformation techniques are well known in the art for most commercial plant species.

In the case of dicotyledonous species, *Agrobacterium*-mediated transformation is generally a preferred technique as it has broad application to many dicotyledons species and is generally very efficient. *Agrobacterium*-mediated transformation generally involves the co-cultivation of *Agrobacterium* with explants from the plant and follows procedures and protocols that are known in the art. Transformed tissue is generally regenerated on medium carrying the appropriate selectable marker. Protocols are known in the art for many dicotyledonous crops including (for example) cotton, tomato, canola and oilseed rape, poplar, potato, sunflower, tobacco and soybean (see for example EP 0 317 511, EP 0 249 432, WO 87/07299, US 5,795,855).

In addition to *Agrobacterium*-mediated transformation, various other techniques can be applied to dicotyledons. These include PEG and electroporation-mediated transformation of protoplasts, and microinjection (see for example Potrykus *et al.* *Mol. Gen. Genet.* 199: 169-177 (1985); Reich *et al.* *Biotechnology* 4: 1001-1004 (1986); Klein *et al.* *Nature* 327: 70-73 (1987)). As with *Agrobacterium*-mediated transformation, transformed tissue is

generally regenerated on medium carrying the appropriate selectable marker using standard techniques known in the art.

Although *Agrobacterium*-mediated transformation has been applied successfully to
5 monocotyledonous species such as rice and maize and protocols for these approaches are available in the art, the most widely used transformation techniques for monocotyledons remain particle bombardment, and PEG and electroporation-mediated transformation of protoplasts.

10 In the case of maize, Gordon-Kamm *et al.* (Plant Cell 2: 603-618 (1990)), Fromm *et al.* (Biotechnology 8: 833-839 (1990) and Koziel *et al.* (Biotechnology 11: 194-200 (1993)) have published techniques for transformation using particle bombardment.

In the case of rice, protoplast-mediated transformation for both *Japonica*- and *Indica*-types
15 has been described (Zhang *et al.* Plant Cell Rep. 7: 379-384 (1988); Shimamoto *et al.* Nature 338: 274-277; Datta *et al.* Biotechnology 8: 736-740 (1990)) and both types are also routinely transformable using particle bombardment (Christou *et al.* Biotechnology 9: 957-962 (1991)).

20 In the case of wheat, transformation by particle bombardment has been described for both type C long-term regenerable callus (Vasil *et al.* Biotechnology 10: 667-674 (1992)) and immature embryos and immature embryo-derived callus (Vasil *et al.* Biotechnology 11: 1553-1558 (1993); Weeks *et al.* Plant Physiol. 102: 1077-1084 (1993)). A further technique is described in published patent applications WO 94/13822 and WO 95/33818.

25

The DNA binding protein constructs of the invention are suitable for expression in a variety of different organisms. However, to enhance the efficiency of expression it may be necessary to modify the nucleotide sequence encoding the DNA binding protein to account for different frequencies of codon usage in different host organisms. Hence it is preferable
30 that the sequences to be introduced into organisms, such as plants, conform to preferred usage of codons in the host organism.

In general, high expression in plants is best achieved from codon sequences that have a GC content of at least 35% and preferably more than 45%. This is thought to be because the existence of ATTTA motifs destabilize messenger RNAs and the existence of AATAAA motifs may cause inappropriate polyadenylation, resulting in truncation of transcription.

5 Murray *et al.* (Nucl. Acids Res. 17: 477-498 (1989)) have shown that even within plants, monocotyledonous and dicotyledonous species have differing preferences for codon usage, with monocotyledonous species generally preferring GC richer sequences. Thus, in order to achieve optimal high level expression in plants, gene sequences can be altered to accommodate such preferences in codon usage in such a manner that the codons encoded
10 by the DNA are not changed.

Plants also have a preference for certain nucleotides adjacent to the ATG encoding the initiating methionine and for most efficient translation, these nucleotides may be modified. To facilitate translation in plant cells, it is preferable to insert, immediately upstream of the
15 ATG representing the initiating methionine of the gene to be expressed, a "plant translational initiation context sequence". A variety of sequences can be inserted at this position. These include the sequence the sequence 5'-AAGGAGATATAACAATG-3' (Prasher *et al.* Gene 111: 229-233 (1992); Chalfie *et al.* Science 263: 802-805 (1992)), the sequence 5'-GTCGACCAATG-3' (Clontech 1993/1994 catalog, page 210), and the
20 sequence 5'-TAAACAATG-3' (Joshi *et al.* Nucl. Acids Res. 15: 6643-6653 (1987)). For any particular plant species, a survey of natural sequences available in any databank (*e.g.* GenBank) can be undertaken to determine preferred "plant translational initiation context sequences" on a species-by-species basis.

25 Any changes that are made to the coding sequence can be made using techniques that are well known in the art and include site directed mutagenesis, PCR, and synthetic gene construction. Such methods are described in published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol) and WO 93/07278 (to Ciba-Geigy). Well known protocols for transient expression in plants can be used to check the expression of
30 modified genes before their transfer to plants by transformation.

C. DNA binding ligands

A DNA binding ligand according to the invention is typically any molecule capable of binding DNA. A variety of DNA binding ligands are known in the art and include acridine orange, 9-Amino-6-chloro-2-methoxyacridine, actinomycin D, 7-aminoactinomycin D, echinomycin, dihydroethidium, ethidium-acridine heterodimer, ethidium bromide, propidium iodide, hexidium iodide, Hoechst 33258, Hoechst 33342, hydroxystibamidine, psoralen, Distamycin A, calicheamicin oligosaccharides, triple-helix forming oligos or PNA, pyrole-imidazole polyamides and peptides or peptide derivatives. These peptides or peptide derivatives are small synthetic polypeptides that can be taken up by plant or animal cells and bind DNA. These polypeptides bind with low affinity to DNA in the absence of a DNA binding molecule but their interaction with DNA may be strengthened by binding of a DNA binding molecule to the target DNA molecule. Such peptide or peptide derivatives have been demonstrated to bind DNA and may be selected from a synthetic library of peptides containing unnatural amino acids as described by Lescrinier et al., Chem. Eur. J. 4:425-433 (1998). Also included within the meaning of the term DNA binding ligand and DNA binding molecules are molecules capable of binding RNA and/or other nucleic acids.

Derivatives of DNA binding ligands are also included provided that they are capable of binding DNA, RNA and/or other nucleic acids.

In a preferred embodiment, a DNA binding ligand according to the invention is capable of modulating the topology, locally or otherwise, of the nucleic acid to which it is bound. In particular, a DNA binding ligand according to the invention may be capable of modulating the topology of a juxtaposed nucleic acid sequence motif to which it is desired to bind a DNA binding molecule according to the invention.

Preferred DNA binding ligands have shape and charge characteristics that allow them to reside along the DNA, in either the minor or major groove, intercalate or a combination of these.

Suitable DNA binding ligands in addition to those known in the art may be selected by the use of nucleic acid binding assays. For example, a candidate DNA binding ligand,

preferably a plurality of candidate DNA binding ligands, is contacted with nucleic acid and binding determined. The nucleic acids may for example be labelled with a detectable label, such as a fluorophore/fluorochrome, such that after a wash step binding can be determined easily, for example by monitoring fluorescence. The nucleic acid with which the candidate
5 binding ligands are contacted may be non-specific nucleic acids, such as a random oligonucleotide library or sonicated genomic DNA and the like. Alternatively, a specific sequence may be used or partially randomised library of sequences.

It is particularly preferred that DNA binding ligands of the invention bind to DNA in a
10 sequence and/or topology dependent manner so that binding can be restricted to a particular target DNA thus enhancing the specificity of the gene switch. Specificity of binding may be determined, for example, by comparing the binding of the DNA binding ligand to a target sequence with binding to a mixture of non-specific DNA molecules.

15 DNA binding ligands according to the invention may bind conditionally to nucleic acid. For example, psoralen is a ligand that can bind DNA covalently if illuminated at wavelengths of about 400 nm or less. Ligands capable of binding nucleic acids in more than one manner may be employed in the current invention. Such ligands may bind or associate with the DNA via any one or more mechanism(s) as outlined above.

20 In a preferred embodiment, libraries of DNA binding ligands may be prepared. In particular, libraries of DNA binding ligands may be immobilised to a solid phase, such as a substantially planar solid phase, including membranes and non-porous substrates such as plastic and glass. The resulting immobilised library may conveniently be used in high
25 throughput screening procedures.

In another preferred embodiment, libraries of synthetic peptides may be prepared. These may be immobilised on a solid phase, such as a bead, and may have weak affinity for DNA. In high throughput screens, DNA target (either specific or a random oligonucleotide) may
30 be labeled with a fluorescent label and the DNA binding molecule may be labeled with an antibody having a different fluorescent label. Interaction of the DNA ligand with DNA may be enhanced in the presence of the DNA binding molecule and the three molecules may be selected by monitoring the fluorescence of the two labels on the solid support.

Particularly preferred DNA binding ligands are those which are substantially non toxic to plants and or animal cells such that they may be administered to said cells and modulate binding of the DNA binding molecule without having an adverse effect on the cells. Thus
5 it may be desirable to pre-screen compounds to exclude toxic compounds.

Furthermore, given that DNA binding ligands should typically be capable of being taken up by the cells of animals or plants, preferred compounds are suitable for administration to animals and plants. For example, preferred compounds are capable of being taken up via
10 the leaves (for foliar application) or roots of plants (for application to the soil) or of permeating seeds (for use in seed treatment). It may also preferred to use compounds that can be taken up by bacteria, yeast and/or fungi that can themselves be delivered to the target host organism. The compounds should also preferably be stable in the soil and/or plant for prolonged periods. In the case of animals, preferred compounds are suitable for
15 topical or oral administration.

D. Target DNA

The term 'target DNA' refers to any DNA for use in the methods of the invention. This
20 DNA may be of known sequence, or may be of unknown sequence. This DNA may be prepared artificially in a laboratory, or may be a naturally occurring DNA. This DNA may be in substantially pure form, or may be in a partially purified form, or may be part of an unpurified or heterogeneous sample. Preferably, the target DNA is a putative promoter or other transcription regulatory region such as an enhancer. More preferably, the target DNA
25 is in substantially pure form. Even more preferably, the target DNA is of known sequence. In a most preferred embodiment, the target DNA is purified DNA of known sequence of a promoter from a gene of interest, for example from a gene suspected of being associated with a disease state, more preferably from a gene useful in gene therapy.

30 Examples of target sequences of interest include sequence motifs that are bound by transcription factors, such as zinc fingers. Particular examples include the promoters of genes involved in the biosynthesis and catabolism of gibberellins (Phillips *et al.*, Plant Physiol 108: 1049-1057 (1995), MacMillin *et al.*, Plant Physiol 113: 1369-1377 (1997),

Williams *et al.*, Plant Physiol 117: 559-563 (1998); Thomas *et al.*, PNAS 96: 4698-4703 (1999)); the promoters of genes whose products are responsible for ripening (such as polygalacturonase and ACC oxidase; the promoters of genes involved in the biosynthesis of volatile ester, which are important flavour compounds in fruits and vegetables (Dudavera *et al.*, Plant Cell 8: 1137-1148 (1996); Dudavera *et al.*, Plant J. 14: 297-304 (1998); Ross *et al.*, Arch. Biochem. Biophys. 367: 9-16 (1999)); the promoters of genes involved in the biosynthesis of pharmaceutically important compounds; and the promoters of genes encoding allergens such as the peanut allergens Arah1, Arah2 and Arah3 (Rabjohn *et al.*, J. Clin. Invest 103: 535-542).

Other plant promoters of interest are the bronze promoter (Ralston *et al.*, Genetics 119: 185-197 (1988) and Genbank Accession No. X07937.1) that directs expression of UDPglucose flavanoid glycosyl-transferase in maize, the patatin-1 gene promoter (Jefferson *et al.*, Plant Mo. Biol. 14: 995-1006 (1990)) that contains sequences capable of directing tuber-specific expression, and the phenylalanine ammonia lyase promoter (Bevan *et al.*, Embo J. 8: 1899-1906 (1989)) though to be involved in responses to mechanical wounding and normal development of the xylem and flower.

Target DNA may also be provided as a plurality of sequences, for example where one or more residues in the nucleic acid sequence are varied or random. Examples of a plurality of sequences are libraries of nucleic acid sequences comprising putative zinc finger binding sites. Other sequence motifs that bind the DNA binding domain of a transcription factor may also be included in the plurality of sequences, typically varied or randomised at one or more positions. For example the chemically inducible promoter fragments described above may be randomised to produce a plurality of target DNA sequences for use in the screening methods of the present invention.

E. Assays

The methods of the present invention typically involve using a tripartite configuration of one or more DNA binding molecules, one or more DNA binding ligands and one or more target DNA sequences as described above to screen for (i) DNA binding molecules that bind to a target DNA in a manner that is modulatable by a DNA binding ligand (ii) DNA

binding ligands that modulate binding of a DNA binding molecule to a target DNA and/or
(iii) a target DNA that is bound modulatably by a DNA binding molecule as a result of an
interaction with a DNA binding ligand. In other words the methods of the invention may
be used to screen for any or all of the components of the gene switch system of the present
invention.

Typically, one or two of the components is a known constant while two or one,
respectively, of the other components are screened. For example, a given DNA binding
molecule and target DNA may be used to screen a plurality of DNA binding ligands or
candidate DNA binding ligands. Alternatively, a plurality of DNA binding molecules and
of DNA binding ligands may be screened against a given target DNA. Other combinations
are also envisaged.

Each component may be one individual molecular species or a plurality of molecular
species. Where a plurality of species is used, they may be substantially all known, partially
randomised or fully randomised. For example, the plurality of DNA binding molecules
may be a randomised zinc finger library and the plurality of target DNA may be a library of
nucleic acid molecules randomised at one or more, typically three or more contiguous,
residues.

However, all three components may be screened for simultaneously. Thus, in a preferred
embodiment, the invention provides a method for isolating multiple DNA binding
molecules in the presence of multiple DNA binding ligands, said DNA binding molecules
being selected using multiple target nucleic acid sequences in a single selection (isolation)
procedure.

The library of candidate DNA binding molecules is preferably a phage display library,
individual candidate molecules of the library optionally being structurally related to zinc
finger transcription factors (for example see Choo and Klug, (1994) PNAS (USA)
91:11163-67, which describes aspects of such libraries and is incorporated herein by
reference). This library is preferably constructed with DNA sequences of the form
GCGNNNGCG (where all 64 middle triplets are represented in the mixture).

One or more DNA binding ligands means at least one DNA binding ligand, preferably two, three or four DNA binding ligands, more preferably five, six, or seven DNA binding ligands, most preferably a mixture of eight DNA binding ligands, or even more. The ligands may be in any molar ratio to one another within the mixture, but will preferably be approximately equimolar with one another.

Said method would preferably be carried out over at least 3, 4, 5 or 6 rounds of selection, preferably about 6 rounds of selection.

DNA binding molecules (such as phage clones) isolated by the above methods would preferably be individually assayed (for example in microtitre plates as described below) for binding to the target DNA (such as a GCGNNNGCG mixture) in the presence and absence of a mixture of the DNA binding ligands to identify clones which are capable of ligand-modulatable binding.

Those phage clones which are capable of ligand-modulatable binding would preferably be tested in the presence of a mixture of the eight ligands, in order to deduce the optimum target DNA sequence, for example using different or variant target DNA sequences, or by the binding site signature method method (see Choo and Klug, (1994) PNAS (USA) 91:11163-67).

Where candidate DNA binding molecules are used rather than molecules known or determined to have DNA binding properties, the method of the invention would preferably feature a pre-selection step to remove candidate DNA binding molecules which do not require ligand to bind the DNA.

Association of the candidate DNA binding molecule with the target DNA may be assessed by any suitable means known to those skilled in the art. For example, the DNA may be immobilised by biotinylation and linking to beads such as streptavidin coated beads (Dynal). In a preferred embodiment wherein the DNA binding molecules are phage displayed polypeptides, binding of said molecules to the DNA may be assessed by eluting those phage which bind, and infecting logarithmic phase *E.coli* TG1 cells. The presence of infective particles eluted from the DNA indicates that association of the DNA binding

molecule(s) with the DNA has occurred. Alternatively, association of the candidate DNA binding molecule(s) with the target DNA may be assessed by Scintillation Proximity Assay (SPA). For example, the target DNA could be biotinylated and immobilised to streptavidin coated SPA beads, and the candidate DNA binding molecules may be radioactively labelled, for example with ³⁵S-Methionine where the molecules are polypeptides. Association of the candidate DNA binding molecules with the target DNA could then be assessed by monitoring the readout of the SPA. Alternatively, the association could be monitored by fluorescent resonance energy transfer (FRET). In this case, the target DNA could be labelled with a donor fluor, and the DNA binding molecule(s) could be labelled with a suitable acceptor fluor. Whilst the two entities are separated, no FRET would be observed, but if association (binding) took place, then there would be a change in the amount of FRET observed, this allowing assessment of the degree of association.

Association of the candidate DNA binding molecule with the target DNA may also be assessed by bandshift assays. Bandshift assays are conducted by measuring the mobility of one or more of the components of the assay, for example the mobility of the DNA, as it is electrophoresed through a suitable gel such as a polyacrylamide acrylamide gel, as is well known to those skilled in the art. In order to assess the association of the candidate DNA binding molecule with the target DNA, the mobility of the DNA could be measured in the presence and absence of the candidate DNA binding molecule. If the mobility of the target DNA is essentially the same in the presence or absence of the candidate DNA binding molecule, then it may be inferred that the molecules do not associate, or that the association is weak. If the mobility of the DNA is retarded in the presence of the candidate DNA binding molecule, then it may be inferred that the candidate molecule is associating with or binding to the DNA.

Association of the candidate DNA binding molecule with the target DNA may also be assessed using filter binding assays. For example, the target DNA molecule may be immobilised on a suitable filter, such as a nitrocellulose filter. The candidate DNA binding molecule may then be labelled, for example radioactively labelled, and contacted with the immobilised target DNA. The binding of or association with the target DNA may be assessed by comparing the amount of labelled candidate DNA binding molecule which associates with the filter only to the amount of labelled candidate DNA binding molecule

which associates with the filter-immobilised target DNA. If more labelled candidate DNA binding molecule associates with the immobilised DNA than with the filter only, it may be inferred that the target DNA molecule does indeed associate with the candidate binding molecule.

5

Binding affinities may be estimated by any suitable means known to those skilled in the art. Binding affinities for the purposes of this invention may be absolute or may be relative. Binding affinities may be determined biochemically, or may simply be estimated by assessing the association of the candidate DNA binding molecule with the target DNA as described above. As used herein, the term binding affinity may refer to a simple estimation of the association of one component of the system with another.

Another suitable detection method is the use of target DNA sequences linked to reporter constructs, such as bacterial luciferase or lacZ. Preferably, the reporter gene product can be measured using optical detection techniques. By way of example, a multiarray format could be used with a different candidate ligand in each position in the array (such as a microtitre plate well) and the same library of zinc finger proteins and target DNA sequences at each position. The zinc finger proteins will generally be fused to a transcriptional activation domain such as the GAL4 acidic activation domain. Transcription may then be compared in the various wells and wells showing a variation in transcription compared to a control well with no ligand may be selected and the ligand further tested to identify specific target sequences/zinc finger proteins whose interaction is affected. These further tests may again be performed using an array format in which this time the DNA binding ligand is kept constant and the target sequence/zinc fingers varied. Phage display techniques as described above may be used to simplify the isolation of suitable zinc finger proteins. Although described in the context of zinc fingers, this method could be applied to other DNA binding molecules.

It is envisaged that the methods of the invention may be applied *in vivo*, for example they could be applied to the selection or isolation of DNA binding molecules capable of associating with target DNA *in vivo* inside one or more cells, in a manner analogous to the one-hybrid system.

It is envisaged that the methods of the invention may be practised in parallel. For example, multiple target DNAs could be used in a single selective step, thereby enabling multiple DNA binding molecules to be isolated simultaneously, even in the same physical vessel. Said multiple DNA binding molecules may preferably be different from one another. Said multiple DNA binding molecules may have similar or identical DNA binding specificities, or may preferably have different DNA binding specificities.

The invention may be worked using multiple DNA binding ligands, either separately or in combination. For example, a target nucleic acid sequence may be used to isolate DNA binding molecules according to the methods essentially as disclosed above, with the modification that more than one DNA binding ligand may be present. In this way, it is possible to isolate multiple DNA binding molecules which require different ligands to bind to the same target nucleic acid sequence(s).

By way of example, a particular embodiment of the method of the invention is as follows:

1. Bacterial colonies containing phage libraries that express a library of zinc fingers randomised at one or more DNA binding residues (see section A.) are transferred from plates to culture medium. Bacterial cultures are grown overnight at 30°C. Culture supernatant containing phages is obtained by centrifugation.

2. 10 pmol of biotinylated target DNA immobilised on 50 mg streptavidin beads (Dynal) is incubated with 1 ml of the bacterial culture supernatant diluted 1:1 with PBS containing 50 μ M ZnCl₂, 4% Marvel, 2% Tween for 1 hour at 20°C on a rolling platform as a preselection step to remove phage that bind to the target DNA in the absence of a ligand.

3. After this time, 0.5 ml of phage solution is transferred to a streptavidin coated tube and incubated with biotinylated DNA target site in the presence of a candidate DNA binding ligand and 4 μ g poly [d(I-C)]. After a one hour incubation the tubes are washed 20 times with PBS containing 50 μ M ZnCl₂ and 1% Tween, and 3 times with PBS containing 50 μ M ZnCl₂ to remove non-binding phage.

4. The remaining phage are eluted using 0.1 ml 0.1 M triethylamine and the solution is neutralised with an equal volume of 1 M Tris-Cl (pH 7.4).

5. Logarithmic-phase *E. coli* TG1 cells are infected with eluted phage, and grown overnight, as described above, to prepare phage supernatants for subsequent rounds of selection.

6. After 4 rounds of selection (steps 1 to 5), bacteria are plated and phage prepared from 96 colonies are screened for binding to the DNA target site in the presence and absence of the ligand. Binding reactions are carried out in wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) and contain 50 μ l of phage solution (bacterial culture supernatant diluted 1:1 with PBS containing 50 μ M ZnCl₂, 4% Marvel, 2% Tween), 0.15 pmol DNA target site and 0.25 μ g poly [d(I-C)]. When added, the DNA binding ligand is present at a concentration of about 1 μ M.

7. After a one hour incubation the wells are washed 20 times with PBS containing 50 μ M ZnCl₂ and 1% Tween (and also ligand at a concentration of 1 μ M where appropriate), and 3 times with PBS containing 50 μ M ZnCl₂.

8. Bound phage are detected by ELISA (carried out in the presence of the ligand at a concentration of about 1 μ M where appropriate) with horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech) and quantitated using SOFTMAX 2.32 (Molecular Devices).

9. Single colonies of transformants obtained after four rounds of selection as described, are grown overnight in culture. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the SequenaseTM 2.0 kit (U.S. Biochemical Corp.). The amino acid sequences of the zinc finger clones are deduced.

In the above example, only one target DNA sequence was used. Where a library of DNA sequences is used, the library of sequences can be screened using the ligand and selected phage expressing the zinc finger of interest to identify specific target DNA sequences. This may conveniently be carried out with the DNA sequences arrayed onto a solid substrate.

In the above example, the zinc fingers (DNA binding molecules) are present on phage. However, alternative methods for displaying the DNA molecules could be used. As described in section A above, an entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, (1994) Proc Natl Acad Sci U S A, 91, 9022-6) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them. Using a library of RNA/ribosomes expressing the DNA binding molecules, screening is

performed in a similar manner to the phage display method except that typically, after an initial preselection step to remove DNA binding molecules that bind in the absence of the ligand only one selection step is performed and the resulting DNA binding molecules identified by cloning the RNA from the RNA/ribosome complexes and sequencing the clones obtained.

To assist in isolating and/or identifying complexes comprising a target DNA, a DNA binding molecule and a DNA binding ligand, it may be desirable to label one or more of the components with a detectable label. For example, the DNA may be labelled with a fluorescent tag and the DNA binding molecule labelled with biotin, such that an enzyme conjugate such as horse radish peroxidase (HRP), that catalyses an optically detectable change in a substrate (different from the fluorescent tag) can be used. If the DNA binding ligand is attached to a bead, then tripartite complexes can be detected because they will both fluoresce and give HRP activity.

A further method which is useful where multiple candidate DNA binding ligands are to be screened involves the use of beads to which are attached different peptide tags. Known combinatorial chemistry techniques are used to produce a library of beads whereby the peptide tag can be used to identify unambiguously the ligand attached to the same bead. Complexes comprising the ligand, a target DNA and a DNA binding molecule can be identified by the use of labelled target DNA and DNA binding molecules as described above. Beads comprising a tripartite complex can then be selected and the identity of the tag determined by spectroscopy techniques which will then give the identity of the ligand.

In general, a bead format is advantageous since it allows easier isolation of productive tripartite complexes and prescreening.

In a further aspect of the invention, DNA binding molecules according to the invention may be advantageously used to determine the sequence composition of a sample of target DNA. For example, a DNA binding molecule according to the invention may be prepared which binds to a known target DNA sequence. By applying this molecule to, or contacting it with, one or more test DNA samples and monitoring its binding thereto, it is possible to determine whether said DNA sample(s) contain the cognate DNA recognition site of the

DNA binding molecule, and therefore derive information about the nucleotide composition of said DNA test sample(s). Such analyses may be advantageously conducted using the binding site signature method (see Choo and Klug, (1994) PNAS (USA) 91:11163-67).

- 5 Individual phage clones could advantageously be assayed for binding of their cognate DNA sequence(s) in the presence or absence of individual ligands, to monitor which particular ligand modulates binding.

Clearly, it may be that more than one ligand modulates binding of DNA binding molecules to their cognate DNA sequence(s). Preferably, individual DNA binding molecules (ie. phage clones) may be assayed for binding to target DNA sequence(s) in the presence of discrete ligand mixtures, wherein each ligand mixture preferably contains a unique mixture of ligands. In this way, the particular ligands which may modulate binding of a particular DNA binding molecule to its cognate target DNA sequence may advantageously be determined. For example, if it is found that two mixtures - one lacking ligand X and the other lacking ligand Y - are incapable of inducing binding, then a mixture of ligands X and Y may have the effect of modulating the binding. This could advantageously be further investigated according to the methods of the invention as described herein.

It is envisaged that this invention may be advantageously used in the isolation of a DNA binding ligand that is capable of modulating the association of a particular DNA binding molecule with its target DNA sequence. Accordingly, the invention provides a method for isolating one or more DNA binding ligands, said ligands each binding one or more target DNA sequence(s), wherein said binding to one or more target DNA sequence(s) modulates the binding of one or more DNA binding molecules, and wherein said DNA binding molecule(s) and said DNA binding ligands are different, said method comprising:

- a) providing one or more target DNA molecule(s);
- b) contacting the target DNA molecule(s) with one or more DNA binding molecule(s)
- 30 c) providing a library of candidate DNA binding ligands,
- d) assessing the ability of candidate DNA binding ligands to modulate the association of the DNA binding molecule(s) with the target DNA molecule(s); and

- e) isolating those candidate DNA binding ligands which modulate the association of the DNA binding molecule(s) with the target DNA molecule(s).

In order to remove DNA binding molecules (for example phage displayed polypeptides) which bind DNA in a ligand-independent manner from a library, a pre-selection step may optionally be performed in the absence of ligand prior to each round of selection. This step removes from the library those clones which do not require ligand for DNA binding. Optionally, candidate molecules selected in this manner may be screened by ELISA for binding to the DNA target in the presence or absence of the ligand(s).

In the above described methods, in order to remove DNA binding molecules (for example phage displayed polypeptides) which bind DNA in a ligand-dependent manner from a library, a pre-selection step may optionally be performed in the presence of ligand prior to each round of selection. This step removes from the library those clones which require ligand for DNA binding. Optionally, candidate molecules selected in this manner may be screened by ELISA for binding to the DNA target in the presence or absence of the ligand(s).

It is envisaged that the methods of the current invention may be advantageously applied to the selection of molecules capable of binding nucleic acids other than DNA, for example RNA. Structural considerations of RNA binding molecules are discussed in Afshar *et al* (Afshar *et al*, 1999: Curr. Op. Biotech. vol 10 pages 59-63). In particular, ligands suitable for use in the methods of the invention as applied to RNA include those ligands described above, or may be selected from aminoglycosides and their derivatives such as paromomycin, neomycin (for examples see Park *et al.*, 1996: J. Am. Chem. Soc. vol 118 pp10150-10155); aminoglycoside mimetics (Tok and Rando 1998: J. Am. Soc. Chem. vol 120 pp 8279-8280); acridine derivatives (for examples see Hamy *et al*, 1998: Biochemistry vol 37 pp5086-5095); small peptides ('aptamers'); polycationic compounds (for examples see Wang *et al*, 1998: Tetrahedron 54 pp7955-7976) or any other nucleic acid binding molecules known to those skilled in the art. In a preferred embodiment, derivatives or libraries of said nucleic acid binding ligands may be prepared.

Accordingly, the present invention provides a method for isolating an RNA binding molecule which binds to a target RNA molecule in a manner modulatable by a RNA-binding ligand, wherein said RNA-binding ligand and said RNA-binding molecule are different, said method comprising; providing a target RNA molecule;

5

(a) contacting the target RNA molecule with a RNA-binding ligand, to produce a RNA-ligand complex;

(b) assessing the ability of candidate RNA-binding molecules to bind the target RNA molecule and the RNA-ligand complex; and isolating those candidate RNA-binding molecules which bind the target RNA molecule and RNA-ligand complex with different binding affinities.

It is further envisaged that the methods of the invention may be advantageously used to select nucleic acid sequences which allow binding of a particular DNA binding ligand/DNA binding molecule combination. For example, one may wish to isolate particular DNA sequences to which a given DNA binding molecule is able to bind, or to isolate only those DNA sequences which depend on the presence of ligand for the DNA binding molecule to associate with them.

Accordingly, there is provided a method for isolating target DNA sequences to which a particular DNA binding molecule will bind, said method comprising

- a) providing a library of target nucleic acid molecule(s);
- b) contacting said nucleic acid molecules with a DNA binding molecule in the presence or absence of DNA binding ligand
- c) assessing the ability of the candidate target DNA molecule(s) to bind the DNA binding molecule; and
- d) isolating those target nucleic acid molecules which bind the DNA binding molecule.

A library of target nucleic acid molecule(s) according to the invention may preferably comprise a plurality of different nucleic acid molecules; preferably said nucleic acid molecules may be related to one another in terms of sequence homology.

A library of candidate nucleic acid binding molecule(s) according to the invention may preferably comprise a plurality of different candidate nucleic acid binding polypeptides; preferably said candidate nucleic acid binding polypeptides may be related to one another in terms of amino acid sequence homology.

5

It is envisaged that this method could be advantageously used in order to isolate DNA sequences which require ligand to associate with a known DNA binding molecule. For example, there may be a DNA sequence which is bound by a known DNA binding molecule in a ligand-independent manner, and it may be desirable to find a DNA
10 sequence(s) which can also associate with the same wild-type DNA binding molecule, but which do so in a ligand-modulatable manner. Preferably, this may be accomplished according to the above method of the present invention.

F. Uses

15

The assay methods of the invention may be used to identify DNA binding molecules, DNA binding ligands and/or target DNA where the binding the DNA binding molecule to the target DNA is modulatable by the DNA binding ligand.

20 These components, such as DNA binding proteins according to the invention and identified by the assay methods of the invention, may be used individually or in combination in a wide variety of applications.

Thus, DNA binding proteins according to the invention and identified by the assay methods
25 of the invention may be employed in a wide variety of applications, including diagnostics and as research tools. Advantageously, they may be employed as diagnostic tools for identifying the presence of particular nucleic acid molecules in a complex mixture. DNA binding molecules according to the invention can preferably differentiate between different target DNA molecules, and their binding affinities for the DNA target sequences are
30 preferably modulated by DNA binding ligand(s). DNA binding molecules according to the invention are useful in switching or modulating gene expression, especially in gene therapy applications and agricultural biotechnology applications as described below.

Specifically, targeted DNA binding molecules, such as zinc fingers, according to the invention may moreover be employed in the regulation of gene transcription, for example by specific cleavage of nucleic acid sequences using a fusion polypeptide comprising a zinc finger targeting domain and a DNA cleavage domain, or by fusion of an transcriptional effector domain to a zinc finger, to activate or repress transcription from a gene which possesses the zinc finger binding sequence in its upstream sequences. Preferably, activation or repression only occurs in the presence of the DNA binding ligand, since in a preferred embodiment the zinc fingers will not bind their target nucleic acid sequences in the absence of the ligand. Alternatively, activation only occurs in the absence of the DNA binding ligand, since the zinc fingers may not bind their target nucleic acid sequences in the presence of the ligand. Zinc fingers capable of differentiating between U and T may be used to preferentially target RNA or DNA, as required. Where RNA-targeting polypeptides are intended, these are included in the term "DNA binding molecule".

Thus DNA binding molecules according to the invention will typically require the presence of a transcriptional effector domain, such as an activation domain or a repressor domain. Examples of transcriptional activation domains include the VP16 and VP64 transactivation domains of Herpes Simplex Virus. Alternative transactivation domains are various and include the maize C1 transactivation domain sequence (Sainz *et al.*, 1997, Mol. Cell. Biol. 17: 115-22) and P1 (Goff *et al.*, 1992, Genes Dev. 6: 864-75; Estruch *et al.*, 1994, Nucleic Acids Res. 22: 3983-89) and a number of other domains that have been reported from plants (see Estruch *et al.*, 1994, *ibid*).

Instead of incorporating a transactivator of gene expression, a repressor of gene expression can be fused to the DNA binding protein and used to down regulate the expression of a gene contiguous or incorporating the DNA binding protein target sequence. Such repressors are known in the art and include, for example, the KRAB-A domain (Moosmann *et al.*, Biol. Chem. 378: 669-677 (1997)) the *engrailed* domain (Han *et al.*, Embo J. 12: 2723-2733 (1993)) and the *snag* domain (Grimes *et al.*, Mol Cell. Biol. 16: 6263-6272 (1996)). These can be used alone or in combination to down-regulate gene expression.

Another possible application is the use of zinc fingers fused to nucleic acid cleavage moieties, such as the catalytic domain of a restriction enzyme, to produce a restriction

enzyme capable of cleaving only target DNA of a specific sequence (see Kim *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160). Using such approaches, different DNA binding domains can be used to create restriction enzymes with any desired recognition nucleotide sequence, but which cleave DNA conditionally dependent on the presence or
5 absence of a particular DNA binding ligand, for instance Distamycin A. It may also be possible to use enzymes other than those that cleave nucleic acids for a variety of purposes.

In a preferred embodiment, the zinc finger polypeptides of the invention may be employed to detect the presence of a particular target nucleic acid sequence in a sample.

10 Accordingly, the invention provides a method for determining the presence of a target nucleic acid molecule, comprising the steps of:

- a) preparing a DNA binding protein by the method set forth above which is specific for
15 the target nucleic acid molecule;
- b) exposing a test system which may comprise the target nucleic acid molecule to the DNA binding protein under conditions which promote binding, and removing any DNA binding protein which remains unbound;
- c) detecting the presence of the DNA binding protein in the test system.

20 Regulation of gene expression *in vivo*

In a particularly preferred embodiment of the present invention, DNA binding molecules capable of binding to a target DNA in a manner modulatable by a DNA binding ligand are
25 used to regulate expression from a gene *in vivo*.

The target gene may be endogenous to the genome of the cell or may be heterologous. However, in either case it will comprise a target DNA sequence, such as a target DNA sequence described above, to which a DNA binding molecule of the invention binds in a
30 manner modulatable by a DNA binding ligand. Where the DNA binding molecule is a polypeptide, it may typically be expressed from a DNA construct present in the host cell comprising the target sequence. The DNA construct is preferably stably integrated into the genome of the host cell, but this is not essential.

Thus in the case of polypeptide DNA binding molecules, a host cell according to the invention comprises a target DNA sequence and a construct capable of directing expression of the DNA binding molecule in the cell.

5

Suitable constructs for expressing the DNA binding molecule are known in the art and are described in section B above. The coding sequence may be expressed constitutively or be regulated. Expression may be ubiquitous or tissue-specific. Suitable regulatory sequences are known in the art and are also described in section B above. Thus the DNA construct
10 will comprise a nucleic acid sequence encoding a DNA binding molecule operably linked to a regulatory sequence capable of directing expression of the DNA binding molecule in a host cell.

It may also be desirable to use target DNA sequences that include operably linked
15 neighbouring sequences that bind transcriptional regulatory proteins, such as transactivators. Preferably the transcriptional regulatory proteins are endogenous to the cell. If not, they typically will need to be introduced into the host cell using suitable nucleic acid constructs.

20 Techniques for introducing nucleic acid constructs into host cells are known in the art for both prokaryotic and eukaryotic cells, including yeast, fungi, plant and animal cells. Many of these techniques are mentioned below in the section on the production of transgenic organisms.

25 Regulation of expression of the gene of interest which comprises a second coding sequence operably linked to the target DNA sequence is typically achieved by administering to the cell a DNA binding ligand according to the invention. Typically, the DNA binding ligand is a molecule such as Distamycin A which may be administered exogenously to the cell and taken up by the cell whereupon it may contact the DNA binding molecule and modulate its
30 binding to the target sequence. However polypeptide DNA binding ligands may also be introduced into the cell either directly or by introducing suitable nucleic acid vectors, including viruses.

The target DNA sequence and the DNA construct encoding the DNA binding molecule are preferably stably integrated into the genome of the host cell. Where the host cell is a single celled organism or part of a multicellular organism, the resulting organism may be termed transgenic. The target DNA may, in a preferred embodiment, be a naturally occurring
5 sequence for which a corresponding DNA binding molecule and DNA binding ligand have been identified using the screening methods of the invention.

The term "multicellular organism" here denotes all multicellular plants, fungi and animals except humans, i.e. prokaryotes and unicellular eukaryotes are excluded specifically. The
10 term also includes an individual organism in all stages of development, including embryonic and fetal stages. A "transgenic" multicellular organisms is any multicellular organism containing cells that bear genetic information received, directly or indirectly, by deliberate genetic manipulation at the subcellular level, such as by microinjection or infection with recombinant virus. Preferably, the organism is transgenic by virtue of
15 comprising at least a heterologous nucleotide sequence encoding a DNA binding molecule or target DNA as herein defined.

"Transgenic" in the present context does not encompass classical crossbreeding or *in vitro* fertilization, but rather denotes organisms in which one or more cells receive a recombinant
20 DNA molecule. Transgenic organisms obtained by subsequent classical crossbreeding or *in vitro* fertilization of one or more transgenic organisms are included within the scope of the term "transgenic".

The term "germline transgenic organism" refers to a transgenic organism in which the
25 genetic information has been taken up and incorporated into a germline cell, therefore conferring the ability to transfer the information to offspring. If such offspring, in fact, possess some or all of that information, then they, too, are transgenic multicellular organisms within the scope of the present invention.

30 The information to be introduced into the organism is preferably foreign to the species of animal to which the recipient belongs (i.e., "heterologous"), but the information may also be foreign only to the particular individual recipient, or genetic information already

possessed by the recipient. In the last case, the introduced gene may be differently expressed than is the native gene.

"Operably linked" refers to polynucleotide sequences which are necessary to effect the expression of coding and non-coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes, generally, such control sequences include promoters and a transcription termination sequence. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

Since the nucleic acid constructs are typically to be integrated into the host genome, it is important to include sequences that will permit expression of polypeptides in a particular genomic context. One possible approach would to use homologous recombination to replace all or part of the endogenous gene whose expression it is desired to regulate with equivalent sequences comprising a target DNA in its regulatory sequences. This should ensure that the gene is subject to the same transcriptional regulatory mechanisms as the endogenous gene, with the exception of the target DNA sequence. Alternatively, homologous recombination may be used in a similar manner but with the regulatory sequences also replaced so that the gene is subject to a different form of regulation.

However, if the construct encoding either the DNA binding molecule or target DNA is placed randomly in the genome, it is possible that the chromatin in that region will be transcriptionally silent and in a condensed state. If this occurs, then the polypeptide will not be expressed – these are termed position-dependent effects. To overcome this problem, it may be desirable to include locus control regions (LCRs) that maintain the intervening chromatin in a transcriptionally competent open conformation. LCRs (also known as scaffold attachment regions (SARs) or matrix attachment regions (MARs)) are well known in the art – an example being the chicken lysozyme A element (Stief *et al.*, 1989, Nature 341: 343), which can be positioned around an expressible gene of interest to effect an increase in overall expression of the gene and diminish position dependent effects upon

incorporation into the organism's genome (Stief *et al.*, 1989, *supra*). Another example is the CD2 gene LCR described by Lang *et al.*, 1991, Nucl. Acid. Res. 19: 5851-5856.

Thus, a polynucleotide construct for use in the present invention, to introduce a nucleotide
5 sequence encoding a DNA binding molecule into the genome of a multicellular organism, typically comprises a nucleotide sequence encoding the DNA binding molecule operably linked to a regulatory sequence capable of directing expression of the coding sequence. In addition the polynucleotide construct may comprise flanking sequences homologous to the host cell organism genome to aid in integration. An alternative approach would be to use
10 viral vectors that are capable of integrating into the host genome, such as retroviruses.

Preferably, a nucleotide construct for use in the present invention further comprises flanking LCRs.

15 Construction of Transgenic Organisms Expressing DNA Binding Molecules

A transgenic organism of the invention is preferably a multicellular eukaryotic organism, such as an animal, a plant or a fungus. Animals include animals of the phyla cnidaria, ctenophora, platyhelminthes, nematoda, annelida, mollusca, chelicerata, uniramia,
20 crustacea and chordata. Uniramians include the subphylum hexapoda that includes insects such as the winged insects. Chordates includes vertebrate groups such as mammals, birds, reptiles and amphibians. Particular examples of mammals include non-human primates, cats, dogs, ungulates such as cows, goats, pigs, sheep and horses and rodents such as mice, rats, gerbils and hamsters.

25

Plants include the seed-bearing plants angiosperms and conifers. Angiosperms include dicotyledons and monocotyledons. Examples of dicotyledonous plants include tobacco, (*Nicotiana plumbaginifolia* and *Nicotiana tabacum*), arabidopsis (*Arabidopsis thaliana*), *Brassica napus*, *Brassica nigra*, *Datura innoxia*, *Vicia narbonensis*, *Vicia faba*, pea (*Pisum sativum*),
30 cauliflower, carnation and lentil (*Lens culinaris*). Examples of monocotyledonous plants include cereals such as wheat, barley, oats and maize.

Production of transgenic animals

Techniques for producing transgenic animals are well known in the art. A useful general textbook on this subject is Houdebine, *Transgenic animals – Generation and Use* (Harwood Academic, 1997) – an extensive review of the techniques used to generate transgenic animals from fish to mice and cows.

Advances in technologies for embryo micromanipulation now permit introduction of heterologous DNA into, for example, fertilized mammalian ova. For instance, totipotent or pluripotent stem cells can be transformed by microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral infection or other means, the transformed cells are then introduced into the embryo, and the embryo then develops into a transgenic animal. In a highly preferred method, developing embryos are infected with a retrovirus containing the desired DNA, and transgenic animals produced from the infected embryo. In a most preferred method, however, the appropriate DNAs are coinjected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos allowed to develop into mature transgenic animals. Those techniques as well known. See reviews of standard laboratory procedures for microinjection of heterologous DNAs into mammalian fertilized ova, including Hogan *et al.*, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Press 1986); Krimpenfort *et al.*, *Bio/Technology* 9:844 (1991); Palmiter *et al.*, *Cell*, 41: 343 (1985); Kraemer *et al.*, *Genetic manipulation of the Mammalian Embryo*, (Cold Spring Harbor Laboratory Press 1985); Hammer *et al.*, *Nature*, 315: 680 (1985); Wagner *et al.*, U.S. Pat. No. 5,175,385; Krimpenfort *et al.*, U.S. Pat. No. 5,175,384, the respective contents of which are incorporated herein by reference

Another method used to produce a transgenic animal involves microinjecting a nucleic acid into pro-nuclear stage eggs by standard methods. Injected eggs are then cultured before transfer into the oviducts of pseudopregnant recipients.

Transgenic animals may also be produced by nuclear transfer technology as described in Schnieke, A.E. *et al.*, 1997, *Science*, 278: 2130 and Cibelli, J.B. *et al.*, 1998, *Science*, 280: 1256. Using this method, fibroblasts from donor animals are stably transfected with a plasmid incorporating the coding sequences for a binding domain or binding partner of

interest under the control of regulatory. Stable transfectants are then fused to enucleated oocytes, cultured and transferred into female recipients.

5 Analysis of animals which may contain transgenic sequences would typically be performed by either PCR or Southern blot analysis following standard methods.

By way of a specific example for the construction of transgenic mammals, such as cows, nucleotide constructs comprising a sequence encoding a DNA binding molecule are microinjected using, for example, the technique described in U.S. Pat. No. 4,873,191, into
10 oocytes which are obtained from ovaries freshly removed from the mammal. The oocytes are aspirated from the follicles and allowed to settle before fertilization with thawed frozen sperm capacitated with heparin and prefractionated by Percoll gradient to isolate the motile fraction.

15 The fertilized oocytes are centrifuged, for example, for eight minutes at 15,000 g to visualize the pronuclei for injection and then cultured from the zygote to morula or blastocyst stage in oviduct tissue-conditioned medium. This medium is prepared by using luminal tissues scraped from oviducts and diluted in culture medium. The zygotes must be placed in the culture medium within two hours following microinjection.

20 Oestrous is then synchronized in the intended recipient mammals, such as cattle, by administering coprostanol. Oestrous is produced within two days and the embryos are transferred to the recipients 5-7 days after estrous. Successful transfer can be evaluated in the offspring by Southern blot.

25 Alternatively, the desired constructs can be introduced into embryonic stem cells (ES cells) and the cells cultured to ensure modification by the transgene. The modified cells are then injected into the blastula embryonic stage and the blastulas replaced into pseudopregnant hosts. The resulting offspring are chimeric with respect to the ES and host cells, and
30 nonchimeric strains which exclusively comprise the ES progeny can be obtained using conventional cross-breeding. This technique is described, for example, in WO91/10741.

Production of transgenic plants

Techniques for producing transgenic plants are well known in the art. Typically, either whole plants, cells or protoplasts may be transformed with a suitable nucleic acid construct encoding a DNA binding molecule or target DNA (see above for examples of nucleic acid constructs). There are many methods for introducing transforming DNA constructs into cells, but not all are suitable for delivering DNA to plant cells. Suitable methods include *Agrobacterium* infection (see, among others, Turpen *et al.*, 1993, J. Virol. Methods, 42: 227-239) or direct delivery of DNA such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles. Acceleration methods are generally preferred and include, for example, microprojectile bombardment. A typical protocol for producing transgenic plants (in particular monocotyledons), taken from U.S. Patent No. 5, 874, 265, is described below.

An example of a method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, non-biological particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming both dicotyledons and monocotyledons, is that neither the isolation of protoplasts nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into plant cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with plant cells cultured in suspension. The screen disperses the tungsten-DNA particles so that they are not delivered to the recipient cells in large aggregates. It is believed that without a screen intervening between the projectile apparatus and the cells to be bombarded, the projectiles aggregate and may be too large for attaining a high frequency of transformation. This may be due to damage inflicted on the recipient cells by projectiles that are too large.

For the bombardment, cells in suspension are preferably concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more clusters of cells transiently expressing a marker gene ("foci") on the bombarded filter. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from 1 to 10 and average 2 to 3.

10 After effecting delivery of exogenous DNA to recipient cells by any of the methods discussed above, a preferred step is to identify the transformed cells for further culturing and plant regeneration. This step may include assaying cultures directly for a screenable trait or by exposing the bombarded cultures to a selective agent or agents.

15 An example of a screenable marker trait is the red pigment produced under the control of the R-locus in maize. This pigment may be detected by culturing cells on a solid support containing nutrient media capable of supporting growth at this stage, incubating the cells at, e.g., 18°C and greater than 180 $\mu\text{E m}^{-2} \text{ s}^{-1}$, and selecting cells from colonies (visible aggregates of cells) that are pigmented. These cells may be cultured further, either in suspension or on solid media.

An exemplary embodiment of methods for identifying transformed cells involves exposing the bombarded cultures to a selective agent, such as a metabolic inhibitor, an antibiotic, herbicide or the like. Cells which have been transformed and have stably integrated a marker gene conferring resistance to the selective agent used, will grow and divide in culture. Sensitive cells will not be amenable to further culturing.

To use the bar-bialaphos selective system, bombarded cells on filters are resuspended in nonselective liquid medium, cultured (e.g. for one to two weeks) and transferred to filters overlaying solid medium containing from 1-3 mg/l bialaphos. While ranges of 1-3 mg/l will typically be preferred, it is proposed that ranges of 0.1-50 mg/l will find utility in the practice of the invention. The type of filter for use in bombardment is not believed to be particularly crucial, and can comprise any solid, porous, inert support.

Cells that survive the exposure to the selective agent may be cultured in media that supports regeneration of plants. Tissue is maintained on a basic media with hormones for about 2-4 weeks, then transferred to media with no hormones. After 2-4 weeks, shoot development will signal the time to transfer to another media.

Regeneration typically requires a progression of media whose composition has been modified to provide the appropriate nutrients and hormonal signals during sequential developmental stages from the transformed callus to the more mature plant. Developing plantlets are transferred to soil, and hardened, e.g., in an environmentally controlled chamber at about 85% relative humidity, 600 ppm CO₂, and 250 $\mu\text{E m}^{-2} \text{ s}^{-1}$ of light. Plants are preferably matured either in a growth chamber or greenhouse. Regeneration will typically take about 3-12 weeks. During regeneration, cells are grown on solid media in tissue culture vessels. An illustrative embodiment of such a vessel is a petri dish. Regenerating plants are preferably grown at about 19°C to 28°C. After the regenerating plants have reached the stage of shoot and root development, they may be transferred to a greenhouse for further growth and testing.

Genomic DNA may be isolated from callus cell lines and plants to determine the presence of the exogenous gene through the use of techniques well known to those skilled in the art such as PCR and/or Southern blotting.

Several techniques exist for inserting the genetic information, the two main principles being direct introduction of the genetic information and introduction of the genetic information by use of a vector system. A review of the general techniques may be found in articles by Potrykus (Annu Rev Plant Physiol Plant Mol Biol [1991] 42:205-225) and Christou (Agro-Food-Industry Hi-Tech March/April 1994 17-27).

Thus, in one aspect, the present invention relates to a vector system which carries a construct encoding a DNA binding molecule or target DNA according to the present invention and which is capable of introducing the construct into the genome of an organism, such as a plant.

The vector system may comprise one vector, but it can comprise at least two vectors. In the case of two vectors, the vector system is normally referred to as a binary vector system. Binary vector systems are described in further detail in Gynheung An et al. (1980), Binary Vectors, *Plant Molecular Biology Manual A3*, 1-19.

5

One extensively employed system for transformation of plant cells with a given promoter or nucleotide sequence or construct is based on the use of a Ti plasmid from *Agrobacterium tumefaciens* or a Ri plasmid from *Agrobacterium rhizogenes* (An et al. (1986), *Plant Physiol.* 81, 301-305 and Butcher D.N. et al. (1980), *Tissue Culture Methods for Plant Pathologists*, eds.: D.S. Ingrams and J.P. Helgeson, 203-208).

10

Several different Ti and Ri plasmids have been constructed which are suitable for the construction of the plant or plant cell constructs described above.

15 Examples of specific applications

The DNA binding molecule/ target DNA/ DNA binding ligand combination may be used to regulate the expression of a nucleotide sequence of interest, such as in a cell of an organism, including prokaryotes, yeasts, fungi, plants and animals, for example mammals, including humans.

20

Nucleotide sequences of interest include genes associated with disease in humans and animals and therapeutic genes. Thus a DNA binding molecule may be used in conjunction with a target DNA sequence and DNA binding ligand in a method of treating or preventing disease in an animal or human patient.

25

Alternatively, a genetic switch of the invention comprising a DNA binding molecule a target DNA sequence and a DNA binding ligand wherein the DNA binding ligand modulates binding of the DNA molecule to the target DNA may be used to regulate expression of a nucleotide sequence of interest in a plant. Examples of specific applications include the following:

30

1. Improvement of ripening characteristics in fruit. A number of genes have been identified that are involved in the ripening process (such as in ethylene biosynthesis). Control of the ripening process via regulation of the expression of those genes will help reduce significant losses via spoilage.

5

2. Modification of plant growth characteristics through intervention in hormonal pathways. Many plant characteristics are controlled by hormones. Regulation of the genes involved in the production of and response to hormones will enable produce crops with altered characteristics.

10

3. Improvement of other characteristics by manipulation of plant gene expression. Overexpression of the Na⁺/H⁺ antiport gene has resulted in enhanced salt tolerance in Arabidopsis. Targetted zinc fingers could be used to regulate the endogenous gene.

15 4. Improvement of plant aroma and flavour. Pathways leading to the production of aroma and flavour compounds in vegetables and fruit are currently being elucidated allowing the enhancement of these traits using gene switch technology.

20 5. Improving the pharmaceutical and nutraceutical potential of plants. Many pharmaceutically active compounds are known to exist in plants, but in many cases production is limited due to insufficient biosynthesis in plants. Gene switch technology could be used to overcome this limitation by upregulating specific genes or biochemical pathways. Other uses include regulating the expression of genes involved in biosynthesis of commercially valuable compounds that are toxic to the development of the plant.

25

6. Reducing harmful plant components. Some plant components lead to adverse allergic reaction when ingested in food. Gene switch technology could be used to overcome this problem by downregulating specific genes responsible for these reactions.

30 7. As well as modulating the expression of endogenous genes, heterologous genes may be introduced whose expression is regulated by a gene switch of the invention. For example, a nucleotide sequence of interest may encode a gene product that is preferentially toxic to cells of the male or female organs of the plant such that the ability of the plant to

reproduce can be regulated. Alternatively, or in addition, the regulatory sequences to which the nucleotide sequence is operably linked may be tissue-specific such that expression when induced only occurs in male or female organs of the plant. Suitable sequences and/or gene products are described in WO89/10396, WO92/04454 (the TA29 promoter from tobacco) and EP-A-344,029, EP-A-412,006 and EP-A-412,911.

Other uses include regulating the expression of genes involved in biosynthesis of commercially valuable compounds that are toxic to the development of the plant.

The present invention will now be described by way of the following examples, which are illustrative only and non-limiting. The examples refer to the figures:

Brief Description of the Figures

Figure 1 shows a graph of the effect of Distamycin A concentration on binding of two different phage (clone 3 (3/2F) and clone 4 (4/5F)) to the DNA sequence AAAAAGGCG. In this case, the small molecule causes phage binding to DNA..

Figure 2 shows a graph of the effect of Actinomycin D concentration on binding of two different phage (AD clone 1 and 6) to the DNA sequence AGCTTGGCG. In this case, the small molecule causes phage binding to DNA..

Figure 3 shows four different phage (0.4/1, 0.4/2, 0.4/4 and 0.4/5) binding to the randomised DNA oligo YRYRYGGCG (where Y is C or T and R is G or A) in the presence, but not in the absence, of echinomycin (EM).

Figure 4 shows the binding site signature of phage 0.4/4 selected using the randomised DNA sequence (Y1)(R2)(Y3)(R4)(Y5)GGCG. The phage has a preference for the DNA sequence (T)(G/A)(C)(G/A)(T) in the presence of echinomycin.

30

Figure 5 shows binding of the phage 0.4/4 to three related DNA sequences, TACGTGGCG, TGTATGGCG and CGTACGGCG, as a function of echinomycin

concentration. The first DNA site contains the optimal binding sequence as revealed by the binding site signature.

Figure 6 shows a graph of the effect of ligand concentration on binding of two different phage to specific DNA sequences. In this case, the respective phage are dissociated from the DNA in the presence of distamycin A or actinomycin D.

Examples

Example 1 - Preparation and Screening of a Zinc Finger Phage Display Library

Selection Of Zinc Finger Phage Binding DNA Targets In The Presence Of Small Molecules

Example 1.1 Selection of Zinc Finger Phage that Bind DNA In The Presence Of Distamycin A

A powerful method of selecting DNA binding proteins is the cloning of peptides (Smith (1985) Science 228, 1315-1317), or protein domains (McCafferty *et al.*, (1990) Nature 348:552-554; Bass *et al.*, (1990) Proteins 8:309-314), as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. A phage display library is created comprising variants of the middle finger from the DNA binding domain of Zif268.

Materials And Methods

Construction And Cloning Of Genes.

In general, procedures and materials are in accordance with guidance given in Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, 1989. The gene for the Zif268 fingers (residues 333-420) is assembled from 8 overlapping synthetic oligonucleotides (see Choo and Klug, (1994) PNAS (USA) 91:11163-67), giving *SfiI* and *NotI* overhangs. The genes for fingers of the phage library are synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which

contain sites for *NotI* and *SfiI* respectively. Backward PCR primers in addition introduce Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these are followed by the residues of the wild type or library fingers as required. Cloning overhangs are produced by digestion with *SfiI* and *NotI* where necessary. Fragments are ligated to 1
5 μ g similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1 (Hoogenboom *et al.*, (1991) Nucleic Acids Res. 19, 4133-4137) in which a section of the pelB leader and a restriction site for the enzyme *SfiI* (underlined) have been added by site-directed mutagenesis using the oligonucleotide:

10 5' CTCCTGCAGTTGGACCTGTGCCATGGCCGGCTGGGCGCATAGAATGG
AACAACTAAAGC 3' (Seq ID No. 1)

which anneals in the region of the polylinker. Electrocompetent DH5 α cells are transformed with recombinant vector in 200ng aliquots, grown for 1 hour in 2xTY medium
15 with 1% glucose, and plated on TYE containing 15 μ g/ml tetracycline and 1% glucose.

The zinc finger phage display library of the present invention contains amino acid randomisations in putative base-contacting positions from the second and third zinc fingers of the three-finger DNA binding domain of Zif268, and contains members that bind DNA
20 of the sequence XXXXXGGCG where X is any base. Further details of the library used may be found in WO 98/53057, which is incorporated herein by reference. The DNA sequences AAAAAAGGCG and AAAAAAGGCGAAAAAA are used as selection targets in this example because short runs of adenines can cause intrinsic DNA bending -
25 moreover, the structure of the bend can be disrupted by binding of the antibiotic distamycin A.

Phage Selection.

Bacterial colonies containing zinc finger phage libraries are transferred from plates to 200ml 2xTY medium (16g/litre Bactotryptone, 10g/litre Bactoyeast extract, 5g/litre NaCl)
30 containing 50 μ M ZnCl₂ and 15 μ g/ml tetracycline. Bacterial cultures are grown overnight at 30°C. Culture supernatant containing phages is obtained by centrifuging at 1500xg for 5 minutes.

Phage selection is over 4 rounds. Before each round, a pre-selection step is included comprising binding of 10 pmol of biotinylated DNA target sites immobilised on 50mg streptavidin coated beads (Dyna) to 1 ml of phage solution (bacterial culture supernatant diluted 1:1 with PBS containing 50 μ M ZnCl₂, 4% Marvel, 2% Tween), for 1 hour at 20°C

5 on a rolling platform. After this time, 0.5 ml of phage solution is transferred to a streptavidin coated tube and incubated with 2 pmol biotinylated DNA target site in the presence of 2 μ M distamycin A (Sigma) and 4 μ g poly [d(I-C)]. After a one hour incubation the tubes are washed 20 times with PBS containing 50 μ M ZnCl₂ and 1% Tween, and 3 times with PBS containing 50 μ M ZnCl₂. Phage are eluted using 0.1ml 0.1M

10 triethylamine and the solution is neutralised with an equal volume of 1M Tris-Cl (pH 7.4). Logarithmic-phase *E. coli* TG1 cells are infected with eluted phage, and grown overnight, as described above, to prepare phage supernatants for subsequent rounds of selection.

After 4 rounds of selection, bacteria are plated and phage prepared from 96 colonies are

15 screened for binding to the DNA target site in the presence and absence of distamycin A. Binding reactions are carried out in wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) and contain 50 μ l of phage solution (bacterial culture supernatant diluted 1:1 with PBS containing 50 μ M ZnCl₂, 4% Marvel, 2% Tween), 0.15 pmol DNA target site and 0.25 μ g poly [d(I-C)]. When added, distamycin A is present at a

20 concentration of 2 μ M. After a one hour incubation the wells are washed 20 times with PBS containing 50 μ M ZnCl₂ and 1% Tween (and also distamycin A at a concentration of 2 μ M where appropriate), and 3 times with PBS containing 50 μ M ZnCl₂. Bound phage are detected by ELISA (carried out in the presence of distamycin A at a concentration of 2 μ M where appropriate) with horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia

25 Biotech) and quantitated using SOFTMAX 2.32 (Molecular Devices).

Sequencing Of Selected Phage.

Single colonies of transformants obtained after four rounds of selection as described, are grown overnight in 2xTY/Zn/Tet. Small aliquots of the cultures are stored in 15% glycerol at -20°C, to be used as an archive. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the Sequenase™ 2.0 kit (U.S. Biochemical Corp.). The amino acid sequences of the zinc finger clones are deduced.

Amino acid sequences from helical regions of zinc fingers selected to bind DNA in the presence of distamycin

10		F1	F2	F3
		- 1 1 2 3 4 5 6	- 1 1 2 3 4 5 6	- 1 1 2 3 4 5 6
15	Clone 1	RSDELTR	RSDDLST	TNNTRIK
	Clone 2	RSDELTR	RSDDLST	HKATRIK
	Clone 3	RSDELTR	RSDDLST	TDKVRKK
20	Clone 4	RSDELTR	RSDDLST	HNASRIN
	Clone 5	RSDELTR	RSDDL SV	TNNSRKK
25	Clone 6	RSDELTR	RSDDLST	TNATRKK
	Clone 7	RSDELTR	RSDDL SQ	TRNTRKN
	Clone 8	RSDELTR	RSDDL SV	TNNSRKN
30				

Clones 1-4 were selected to bind the oligo:
tataAAAAAAGGCGTGt**cacagtcagtcacacgtc**

35 Clones 5-8 were selected to bind the oligo:
tataAAAAAAGGCGAAAAA**tacagtcagtcacacgtc**

Zinc finger phage clones are isolated according to this method which bind the target with higher affinity in the presence of ligand than in the absence of ligand (see Figure 1). This

method also selected certain clones that bound DNA in the absence of the ligand but were displaced from the DNA in the presence of the ligand (see Example 1.4 below).

Example 1.2 - Selection of Zinc Finger Phage Binding DNA In The Presence of Actinomycin D

An adaptation to the method outlined in the Example 1.1 was used to isolate phage that bound DNA in the presence of a different small molecule, actinomycin D. In this example the DNA target was AGCTTGGCG.

Phage Selection

Essentially the method was the same as used in the previous section using four rounds of a preselection step followed by a selection step, washing and elution. Differences in the method are described. The preselection step comprised of 7.5 pmol of biotinylated DNA target site immobilised on 18.75 µl streptavidin coated beads (Dynal) in a 100 µl mixture containing 4 µl phage library 96 µl PBS, 2% Marvel, 1% Tween-20, 50 µM ZnCl₂ for 1 hour at room temperature with constant mixing. Phage selections were made in streptavidin coated tubes with the phage supernatant, 5 nM biotinylated target DNA, 10 µM actinomycin D in the presence of 1 µg poly [d(I-C)] competitor. The selections were incubated for 1 hour at room temperature. The bound phage were washed and eluted as described above.

ELISA was performed as described above but using 5 nM biotinylated target DNA, 0.25 µg poly[d(I-C)] competitor in the assay and 10 µM actinomycin D where appropriate. Phage were sequenced using Big Dye Terminator Cycle Sequencing Kit (Perkin Elmer Biosystems) and automated sequencing.

The amino acid sequences from the helical regions of the selected zinc fingers were sequenced as:

clone 1	RSDELTRHIRIH	RSDTLSVHIRTH	HNAHRKTHTKIH
clone 6	RSDELTRHIRIH	RSDHLSVHIRTH	KKFAHSAHRKTHTKIH

These two clones were selected using the oligo:

tatacaAGCTTGGCGatcacagtcagtccacagtc

- 5 These zinc finger clones bind to the target oligo with higher affinity in the presence of actinomycin D than in the absence of DNA binding ligand (see Figure 2).

Example 1.3 - Selection of Zinc Finger Phage Using Randomised DNA In The Presence Of Echinomycin. And Subsequent Deconvolution of Binding Partners

10

In this experiment the library of DNA binding molecules was sorted using a library of DNA sequences in the presence of a small molecule. After DNA binding molecules that bound to DNAs in the presence of the small molecule had been selected, the optimal binding site(s) for each DNA binding molecule were determined using the binding site signature.

15

a) Selections

In this experiment, 50 pmol of DNA target library of sequence YRYRYGGCG (where Y is C or T and R is G or A) was bound to 125 µl of streptavidin coated beads (Dyna) and the beads were used to preselect 0.4 µl of phage library in 100 µl of PBS, 2% Marvel, 1% Tween-20, 50 µM ZnCl₂ for 1 hour at room temperature with constant mixing. Phage selections were made in streptavidin coated tubes with the phage supernatant, 30 nM biotinylated target DNA, 10 µM echinomycin in the presence of 1 µg poly [d(I-C)] competitor. The selections were incubated for 1 hour at room temperature. The bound phage were washed and eluted as described above.

25

ELISA was performed as described above but using 30 nM biotinylated target DNA, 0.5 µg poly[d(I-C)] competitor in the assay and 10 µM echinomycin where appropriate. Phage were sequenced using Big Dye Terminator Cycle Sequencing Kit (Perkin Elmer Biosystems) and automated sequencing.

30

Four different clones were selected using the DNA library tatagtYRYRYGGCG atcacagtcagtccacagtc in the presence of echinomycin (see Figure 3).

The amino acid sequences from the helical regions of the selected zinc fingers were sequenced as:

	clone 0.4/1	RSDELTRHIRIH	RSDHLSKHIRTH	KKFARSQTRINHTKIH
5	clone 0.4/2	RSDELTRHIRIH	RSDHLSEHIRTH	TRNARTKHTKIH
	clone 0.4/4	RSDELTRHIRIH	RSDHLSNHIRTH	RNDTRKTHTKIH
	clone 0.4/5	RSDELTRHIRIH	RSDNLSTHIRTH	KKFAHSNTRKNHTKIH

b) *Binding site signature*

10

The signature of the clone 0.4/4 was determined using a modified binding site signature assay. For each of the 5 randomised positions of the oligo, a base was fixed at one of the five positions whilst the remaining 4 positions contained defined mixtures of bases. For the pyrimidine position the base was fixed as either C or T and for the purine position the base was fixed as either G or A so that by testing each position in turn an optimal sequence or binding site signature could be determined.

15

In each well of a streptavidin-coated microtitre plate 2 µl of phage solution (overnight *E. coli* culture supernatant containing phage) were mixed with 48 µl of 2% Marvel, 1% Tween-20, 0.5 µg poly [d(I-C)], 10 µM echinomycin and between 8-16 nM of biotinylated target DNA. The reaction was incubated for 1 hour at room temperature, followed by 6 washes with PBS containing 1% Tween-20, 50 µM ZnCl₂ and 3 washes with PBS containing 0.05% Tween-20, 50 µM ZnCl₂. 100 µl of PBS containing 1% Marvel, 0.05% Tween-20, 50 µM ZnCl₂ and 1/5000 dilution of anti-M13 horse radish peroxidase antibody conjugate (Amersham Pharmacia Biotech) was added to each well and incubated for 1 hour at room temperature. The ELISA plate was washed 3 times with PBS containing 0.05% Tween-20, 50 µM ZnCl₂ followed by three washes with 3 washes of PBS containing 50 µM ZnCl₂. The assay was developed with BCIP/NBT substrates and quantified using a plate reader.

25

30

This method determined the binding site sequence of clone 0.4/4 to be (T₁)(G/A₂)(C₃)(G/A₄)(T₅) (see Figure 4).

c) *Verification of the target DNA sequence*

The optimal target DNA sequence, as determined by the binding site signature, was synthesised together with two other related DNA sequences that were present in the original random DNA library but differed in some of the optimal base positions of the binding site.

These oligonucleotides had the sequence:

tatagtTACGTGGCGatcacagtcagtcacacgtc

tatagtTGTATGGCGatcacagtcagtcacacgtc

tatagtCGTACGGCGatcacagtcagtcacacgtc

Binding of the phage clone was tested as a function of DNA concentrations (from 5 nM to 0.312 nM) in the presence of 10 µM echinomycin. A phage ELISA was set up using 20 µl phage supernatant, 0.5 µg poly[d(I-C)], 10 µM echinomycin in PBS containing 1% Marvel, 1% Tween-20, 50 µM ZnCl₂. The total volume of the assay was 50 µl. The assay was washed and developed as described as for the binding site signature assay.

This method showed that the clone 0.4/4 bound preferentially to the sequence determined from the binding site signature, i.e. TACGTGGCG, in the presence of the small molecule (see Figure 5).

Example 1.4 Selection of Zinc Finger Phage that are dissociated from their DNA Targets In The Presence of Distamycin A or Actinomycin D

This example describes phage that bound DNA targets with higher affinity in the absence of ligand. These phage were isolated using either: (a) the same method as in example 1.1, or (b) by selection in the absence of small molecule and phage elution from DNA using a small molecule.

In this latter case (b) the method was as follows.

Phage selection is over 4 rounds. Binding reactions contain 10 pmol biotinylated DNA site immobilised on 50mg streptavidin coated beads (Dynal) and a 1 ml solution of zinc finger

phage library (as described in 1.1) Reactions were incubated for 1 h on a rolling platform. After this time, beads were washed 20 times as described in 1.1 and finally phage were eluted from the beads over 5 minutes using a solution containing ligand (10 μ M Distamycin A, or 1 μ M Actinomycin D in PBS/Zn).

5

Some phage isolated by either of the above methods (a or b) bound DNA in the absence of ligand but could be displaced by concentrations of distamycin A at 10 μ M and actinomycin D at 1 μ M. The distamycin sensitive clone was selected using the DNA target AAAAAGCGGAAAAA and its helices were sequenced as:

10

QSRSLIQ QRDSLRS RSDERKR

The actinomycin D sensitive clone was selected with the DNA target AGCTTGGCG and its helices were sequenced as:

15

RSDELTR RSDVLST TRSSRKK

Figure 6 demonstrates the sensitivity of each clone to the respective drug.

20 **Example 2 - Modulation Of Binding Of Polypeptides To Target DNA By DNA Binding Ligand**

Individual phage clones are assayed for modulation of target DNA binding by ligand in a phage ELISA binding assay.

25

Binding assay reactions are carried out in wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) as in Example 1, except that the distamycin concentration is varied while the DNA concentration is kept constant at 2 nM.

30 Induction of higher affinity DNA binding is observed when distamycin is added to the binding reaction at 10^{-6} M - 10^{-7} M.

Binding of the zinc finger phage to DNA in the absence of ligand, or at ligand concentrations of 10^{-9} M or lower, results in phage retention close to background level, i.e. lower affinity binding than in the presence of ligand.

- 5 Background level affinity binding is defined as the phage retention in binding reactions that contain no DNA binding site.

Example 3 - DNA-Ligand Modulatable Restriction Enzyme

- 10 Phage-selected or rationally designed zinc finger domains which bind target DNA sequences in a manner modulatable by a DNA binding ligand can be converted to restriction enzymes which cleave DNA containing said target sequences in a manner modulatable by DNA binding ligand. This is achieved by coupling an appropriate zinc finger, as isolated in Example 1 above, to a cleavage domain of a restriction enzyme or
15 other nucleic acid cleaving moiety.

A method of converting zinc finger DNA binding domains to chimaeric restriction endonucleases has been described in Kim, *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160. In order to demonstrate the applicability of DNA ligand-modulatable zinc
20 fingers to restriction enzymes, a fusion is made between the catalytic domain of Fok I as described by Kim *et al.* and a zinc finger of Example 1. Fusion of the zinc finger nucleic acid-binding domain to the catalytic domain of Fok I restriction enzyme results in a novel endonuclease which cleaves DNA adjacent to the DNA recognition sequence of the zinc finger (AAAAAAGGCG or AAAAAAGGCGAAAAAA).

- 25 The oligonucleotides AAAAAAGGCG and AAAAAAGGCGAAAAAA are synthesised and ligated to arbitrary DNA sequences. After incubation with the zinc finger restriction enzyme, the nucleic acids are analysed by gel electrophoresis. Bands indicating cleavage of the nucleic acid at a position corresponding to the location of the oligonucleotide(s)
30 (AAAAAAGGCG / AAAAAAGGCGAAAAAA) are visible.

In a further experiment, the zinc finger is fused to an amino terminal copper/nickel binding motif. Under the correct redox conditions (Nagaoka, M., *et al.*, (1994) J. Am. Chem.

Soc. 116:4085-4086), sequence-specific DNA cleavage is observed, only in the presence of DNA incorporating oligonucleotide AAAAAAGGCG or AAAAAAGGCGAAAAAA.

Example 4 - Modulation Of Transcriptional Activity *In Vivo*

A reporter system is produced which produces a reporter signal conditionally depending on the binding of the zinc finger DNA binding molecule to its target DNA sequence. This binding, and hence transcription from the reporter system, is modulated by the DNA binding ligand Distamycin A.

A transient transfection system using zinc finger transcription factors is produced as described in Choo, Y., *et al.*, (1997) J. Mol. Biol 273:525-532. This system comprises an expression plasmid which produces a phage-selected zinc finger fused to the activation domain of HSV VP16, and a reporter plasmid which contains the recognition sequence of the zinc finger upstream of a CAT reporter gene.

Thus, a zinc finger which recognises the DNA sequence AAAAAAGGCG is selected by phage display as described in Example 1. By the method of the preceding examples, said zinc finger is used to construct transcription factors as described above.

A transient expression experiment is conducted, wherein the CAT reporter gene on the reporter plasmid is placed downstream of the sequence AAAAAAGGCG. The reporter plasmid is cotransfected with a plasmid vector expressing the zinc finger-HSV fusion under the control of a constitutive promoter. No activation of CAT gene expression is observed.

However, when the same experiment is conducted in the presence of Distamycin A, CAT expression is observed as a result of the binding of the zinc finger transcription factor to its recognition sequence AAAAAAGGCG.

Example 5 - Isolation of cognate target nucleic acids

Using a known DNA binding molecule, target DNA sequences to which it can bind are isolated.

5

The 434 repressor is a gene regulatory protein of phage 434. It binds to a 14bp operator site (see Koudelka *et al.*, 1987, Nature vol 326 pp 886-888). This operator site consists of five conserved bp (1-5), then four variable bp (6-9), then five more conserved bp (10-14) as shown below:

10

Site: 1 5 6 7 8 9 10 14
Base: A C A A G/T X X X X A/T T T G T
wherein X is any base.

15 The conserved bases contact the 434 repressor protein. The four variable bases are thought not to contact the 434 repressor protein. However, the four bases which do not contact the 434 repressor protein may affect the affinity of binding of the repressor to the operator site.

20 The 434 repressor protein (ie. the DNA binding molecule) is contacted with a library of different target DNA sequences in the presence and absence of ligand:

The target DNA sequences are synthesized using an Applied Biosystems 380A DNA synthesizer and are purified by gel electrophoresis. The four variable bases ('X' as shown above) are randomised, producing a library of 256 different target DNA molecules, position 5 being T, and position 10 being A. At the 5' and 3' ends of this sequence are placed PCR primer sequences for amplification and recovery of the central target sequences.

Structure of target DNA sequence library:

5' 1 6 9 14 3'
 GTCGGATCCTGTCTGAGGTGAGACAATXXXXATTGTGTCTTCCGACGTCGAATTCGCG
 wherein X is any base, and the partially randomised 434 operator is underlined.

The 434 repressor protein is added to the library of target DNA sequences, in the presence and absence of 2 μ M distamycin A (Sigma) ligand in 200 μ l binding buffer (9 mM Tris-HCl pH 8.0, 90 mM KCl, 90 μ M ZnSO₄) and incubated for 30 min.

5

Nitrocellulose filters (BA 85, Schleicher and Schüll) are placed into a suction chamber (as in Thiesen *et al. (eds)*, Immunological Methods vol IV, Academic Press, Orlando) and prewet with 600 ml Tris-HCl binding buffer. The protein-oligonucleotide mix is applied to the filter(s) with gentle suction, the filters are washed with 4 ml Tris-HCl binding buffer.

10 Oligonucleotides are eluted in 200 μ l binding buffer plus 1 mM 1-10-o-phenanthroline.

Oligonucleotides are then amplified by PCR, using the following primers:

Primer A 5'-GTCGGATCCTGTCTGAGGTGAG-3'

15 Primer B 5'-CGCGAATTCGACGTCGGAAGAC-3'

using an amplification kit (Perkin Elmer Cetus) with the following cycling regime:

93°C 30 sec; 45°C 120 sec; 45°C to 67°C ramp 60 sec; 67°C 180 sec for 25 cycles.

1 μ l of eluted oligonucleotide material is used as template.

20

Optionally, the PCR amplified DNA product is then used in further rounds of incubation with the 434 repressor protein, nitrocellulose filter binding, oligonucleotide elution and PCR amplification.

25 PCR amplified DNA products are then sequenced using standard techniques.

Target DNA sequences are selected which bind the 434 repressor with higher affinity in the presence of ligand than in the absence of ligand. Furthermore, DNA sequences are selected which bind the 434 repressor in the absence of ligand with a higher affinity than in the

30

presence of ligand.

Example 6 - Isolation of ligands which affect the binding of a DNA binding molecule to its cognate DNA target

5 The 434 repressor protein of Example 5 is used in conjunction with a target operator DNA sequence to which it binds.

The operator sequence used is

5'- A C A A T A A A T A T T G T -3'

10 A library of DNA binding ligands is used in place of the 2 μ M distamycin A (Sigma) DNA binding ligand of Example 5.

Ligands are isolated which are capable of increasing the affinity of the 434 repressor for its cognate DNA target sequence. Ligands are also isolated which are capable of decreasing
15 the affinity of the 434 repressor for its cognate DNA target sequence.

Example 7 - Generation of Transgenic Plants Expressing a Zinc Finger Protein Fused to a Transactivation Domain

20 To investigate the utility of heterologous zinc finger proteins for the regulation of plant genes, a synthetic zinc finger protein was designed and introduced into transgenic *Arabidopsis thaliana* under the control of a promoter capable of expression in a plant as described below. A second construct comprising the zinc finger protein binding sequence fused upstream of the Green Fluorescent Protein (GFP) reporter gene was also introduced
25 into transgenic *Arabidopsis thaliana* as described in Example 8. Crossing the two transgenic lines produced progeny plants carrying both constructs in which the GFP reporter gene was expressed demonstrating transactivation of the gene by the zinc finger protein.

30 Using conventional cloning techniques, the following constructs were made as XbaI-BamHI fragments in the cloning vector pcDNA3.1 (Invitrogen).

pTFIIIAZifVP16

pTFIIIAZifVP16 comprises a fusion of four finger domains of the zinc finger protein TFIIIA fused to the three fingers of the zinc finger protein Zif268. The TFIIIA-derived sequence is fused in frame to the translational initiation sequence ATG. The 7 amino acid Nuclear Localization Sequence (NLS) of the wild-type Simian Virus 40 Large T-Antigen is fused to the 3' end of the Zif268 sequence, and the VP16 transactivation sequence is fused downstream of the NLS. In addition, 30 bp sequence from the *c-myc* gene is introduced downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the transgene. While this is experimentally useful, the presence of this tag is not required for the activation (or repression) of gene expression via zinc finger proteins.

The sequence of pTFIIIAZifVP16 is shown in SEQ ID No. 1 as an XbaI-BamHI fragment. The translational initiating ATG is located at position 15 and is double underlined. Fingers 1 to 4 of TFIIIA extend from position 18 to position 416. Finger 4 (positions 308–416) does not bind DNA within the target sequence, but instead serves to separate the first three fingers of TFIIIA from Zif268 which is located at positions 417–689. The NLS is located at positions 701–722, the VP16 transactivation domain from positions 723–956, and the *c-myc* tag from positions 957–986. This is followed by the translational terminator TAA.

pTFIIIAZifVP64

pTFIIIAZifVP64 is similar to pTFIIIAZifVP16 except that the VP64 transactivation sequence replaces the VP16 sequence of pTFIIIAZifVP16.

The sequence of pTFIIIAZifVP64 is shown in SEQ ID No. 2 as an XbaI-BamHI fragment. Locations within this sequence are as for pTFIIIAZifVP16 except that the VP64 domain is located at position 723–908 and the *c-myc* tag from positions 909–938.

Using conventional cloning techniques, the sequence 5'-AAGGAGATATAACA-3' is introduced upstream of the translational initiating ATG of both pTFIIIAZifVP16 and pTFIIIAZifVP64. This sequence incorporates a plant translational initiation context

sequence to facilitate translation in plant cells (Prasher *et al.* Gene 111: 229-233 (1992); Chalfie *et al.* Science 263: 802-805 (1992)).

The final constructs are transferred to the plant binary vector pBIN121 between the
5 Cauliflower Mosaic Virus 35S promoter and the nopaline synthase terminator sequence. This transfer is effected using the XbaI site of pBIN121. The binary constructs thus derived are then introduced into *Agrobacterium tumefaciens* (strain LBA 4044 or GV 3101) either by triparental mating or direct transformation.

10 Next, *Arabidopsis thaliana* are transformed with *Agrobacterium* containing the binary vector construct using conventional transformation techniques. For example, using vacuum infiltration (*e.g.* Bechtold *et al.* CR Acad Sci Paris 316: 1194-1199; Bent *et al.* Science 265: 1856-1860 (1994)), transformation can be undertaken essentially as follows. Seeds of *Arabidopsis* are planted on top of cheesecloth covered soil and allowed to grow at
15 a final density of 1 per square inch under conditions of 16 hours light/8 hours dark. After 4-6 weeks, plants are ready to infiltrate. An overnight liquid culture of *Agrobacterium* carrying the appropriate construct is grown up at 28°C and used to inoculate a fresh 500ml culture. This culture is grown to an OD₆₀₀ of at least 2.0, after which the cells are harvested by centrifugation and resuspended in 1 litre of infiltration medium (1 litre
20 prepared to contain: 2.2 g MS Salts, 1 X B5 vitamins, 50 g sucrose, 0.5 g MES pH 5.7, 0.044 µM benzylaminopurine, 200 L Silwet µL-77 (OSI Specialty)). To vacuum infiltrate, pots are inverted into the infiltration medium and placed into a vacuum oven at room temperature. Infiltration is allowed to proceed for 5 mins at 400mm Hg. After releasing the vacuum, the pot is removed and layed it on its side and covered with Saran wrap. The
25 cover is removed the next day and the plant stood upright. Seeds harvested from infiltrated plants are surface sterilized and selected on appropriate medium. Vernalizalizon is undertaken for two nights at around 4°C. Plates are then transferred to a plant growth chamber. After about 7 days, transformants are visible and are transferred to soil and grown to maturity.

Many transgenic plants are grown to maturity. They appear phenotypically normal and are selfed to homozygosity using standard techniques involving crossing and germination of progeny on appropriate concentration of antibiotic.

- 5 Transgenic plant lines carrying the TFIIIAZifVP16 construct are designated *At*-TFIIIAZifVP16 and transgenic plant lines carrying the TFIIIAZifVP64 construct are designated *At*-TFIIIAZifVP64.

10 **Example 8 - Generation of Transgenic Plants Carrying a Green Fluorescent Protein Reporter Gene**

A reporter plasmid is constructed which incorporates the target DNA sequence of the TFIIIAZifVP16 and TFIIIAZifVP64 zinc finger proteins described above upstream of the Green Fluorescent Protein (GFP) reporter gene. The target DNA sequence of
15 TFIIIAZifVP16 and TFIIIAZifVP64 is shown in SEQ I.D. No. 3. This sequence is incorporated in single copy immediately upstream of the CaMV 35S -90 minimal promoter to which the GFP gene is fused.

The resultant plasmid, designated pTFIIIAZif-UAS/GFP, is transferred to the plant binary
20 vector pBIN121 replacing the Cauliflower Mosaic Virus 35S promoter. This construct is then transferred to *Agrobacterium tumefaciens* and subsequently transferred to *Arabidopsis thaliana* as described above. Transgenic plants carrying the construct are designated *At*-TFIIIAZif-UAS/GFP.

25 **Example 9 - Use of Zinc Finger Proteins to Up-Regulate a Transgene in a Plant**

To assess whether the zinc finger constructs TFIIIAZifVP16 and TFIIIAZifVP64 are able to transactivate gene expression *in planta*, *Arabidopsis* lines *At*-TFIIIAZifVP16 and *At*-TFIIIAZifVP64 are crossed to *At*-TFIIIAZif-UAS/GFP. The progeny of such crosses
30 yield plants that carry the reporter construct TFIIIAZif-UAS/GFP together with either the zinc finger protein construct TFIIIAZifVP16 or the zinc finger construct TFIIIAZifVP64.

Plants are screened for GFP expression using an inverted fluorescence microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425, dichronic 455, emission LP 460) suitable for the main 395 nm excitation and 509 nm emission peaks of GFP.

- 5 In each case, the zinc finger construct is able to transactivate gene expression demonstrating the utility of heterologous zinc finger proteins for the regulation of plant genes.

Example 10 – Generation of Transgenic Plants Expressing a Zinc Finger Fused to a 10 Plant Transactivation domain

The constructs pTFIIIAZifVP16 and pTFIIIAZifVP64 utilize the VP16 and VP64 transactivation domains of Herpes Simplex Virus to activate gene expression. Alternative transactivation domains are various and include the C1 transactivation domain sequence
15 (from maize; see Goff *et al.*; Genes Dev. 5: 298-309 (1991); Goff *et al.*; Genes Dev. 6: 864-875 (1992)), and a number of other domains that have been reported from plants (see Estruch *et al.*; Nucl. Acids Res. 22: 3983-3989 (1994)).

Construct pTFIIAZifC1 is made as described above for pTFIIIAZifVP16 and
20 pTFIIIAZifVP64 except the VP16/VP64 activation domains are replaced with the C1 transactivation domain sequence

A transgenic *Arabidopsis* line, designated *At*-TFIIAZifC1, is produced as described above in Example 8 and crossed with *At*-TFIIIAZif-UAS/GFP. The progeny of such crosses yield
25 plants that carry the reporter construct TFIIIAZif-UAS/GFP together with either the zinc finger protein construct TFIIIAZifC1.

Plants are screened for GFP expression using an inverted fluorescence microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425, dichronic 455, emission LP
30 460) suitable for the main 395 nm excitation and 509 nm emission peaks of GFP.

Example 11 – Regulation of an endogenous plant gene – UDP glucose flavonoid glucosyl-transferase (UFGT).

To determine whether a suitably configured zinc finger could be used to regulate gene transcription from an endogenous gene in a plant, the maize UDP glucose flavonoid glucosyl-transferase (UFGT) gene (the *Bronze1* gene) was selected as the target gene. UFGT is involved in anthocyanin biosynthesis. A number of wild type alleles have been identified including Bz-W22 that conditions a purple phenotypes in the maize seed and plant. The Bronze locus has been the subject of extensive genetic research because its phenotype is easy to score and its expression is tissue specific and varied (for example aleurone, anthers, husks, cob and roots). The complete sequence of Bz-W22 including upstream regulatory sequences has been determined (Ralston *et al.*, Genetics 119: 185-197). A number of sequence motifs that bind transcriptional regulatory proteins have been identified within the Bronze promoter including sequences homologous to consensus binding sites for the myb- and myc-like proteins (Roth *et al.*, Plant Cell 3: 317-325).

Identification of a zinc finger that binds to the bronze promoter

The first step is to carry out a screen for zinc finger proteins that bind to a selected region of the Bronze promoter. A region is chosen just upstream of the AT rich block located at between -88 and -80, which has been shown to be critical for Bz1 expression (Roth *et al.*, supra).

1. Bacterial colonies containing phage libraries that express a library of zinc fingers randomised at one or more DNA binding residues (see Example 1) are transferred from plates to culture medium. Bacterial cultures are grown overnight at 30°C. Culture supernatant containing phages is obtained by centrifugation.
2. 10 pmol of biotinylated target DNA, derived from the Bronze promoter, immobilised on 50 mg streptavidin beads (Dynal) is incubated with 1 ml of the bacterial culture supernatant diluted 1:1 with PBS containing 50 µM ZnCl₂, 4% Marvel, 2% Tween in a streptavidin coated tube for 1 hour at 20°C on a rolling platform in the presence of 4 µg poly [d(I-C)] as competitor.

3. The tubes are washed 20 times with PBS containing 50 μ M ZnCl₂ and 1% Tween, and 3 times with PBS containing 50 μ M ZnCl₂ to remove non-binding phage.

4. The remaining phage are eluted using 0.1 ml 0.1 M triethylamine and the solution is neutralised with an equal volume of 1 M Tris-Cl (pH 7.4).

5. Logarithmic-phase *E. coli* TG1 cells are infected with eluted phage, and grown overnight, as described above, to prepare phage supernatants for subsequent rounds of selection.

6. Single colonies of transformants obtained after four rounds of selection (steps 1 to 5) as described, are grown overnight in culture. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the SequenaseTM 2.0 kit (U.S. Biochemical Corp.). The amino acid sequences of the zinc finger clones are deduced.

Construction of a vector for expression of the zinc finger clone fused to a C1 activation domain in maize protoplasts

Using conventional cloning techniques and in a similar manner to Example 7, the construct pZifBz23C1 is made in cloning vector pcDNA3.1 (Invitrogen).

pZifBz23C1 comprises a the three fingers of the zinc finger protein clone ZifBz23 fused in frame to the translational initiation sequence ATG. The 7 amino acid Nuclear Localization Sequence (NLS) of the wild-type Simian Virus 40 Large T-Antigen is fused to the 3' end of the ZifBz23 sequence, and the C1 transactivation sequence is fused downstream of the NLS. In addition, 30 bp sequence from the *c-myc* gene is introduced downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the transgene.

The coding sequences of pZifBz23C1 are transferred to a plant expression vector suitable for use in maize protoplasts, the coding sequence being under the control of a constitutive CaMV 35S promoter. The resulting plasmid is termed pTMBz23. The vector also contains a hygromycin resistance gene for selection purposes.

A suspension culture of maize cells is prepared from calli derived from embryos obtained from inbred W22 maize stocks grown to flowering in a greenhouse and self pollinated using essentially the protocol described in EP-A-332104 (Examples 40 and 41). The

suspension culture is then used to prepare protoplasts using essentially the protocol described in EP-A-332104 (Example 42).

Protoplasts are resuspended in 0.2 M mannitol, 0.1% w/v MES, 72 mM NaCl, 70 mM CaCl₂, 2.5 mM KCl, 2.5 mM glucose pH to 5.8 with KOH, at a density of about 2×10^6 per ml. 1 ml of the protoplast suspension is then aliquotted into plastic electroporation cuvettes and 10 µg of linearized pTMBz23 added. Electroporation is carried out as described in EP-A-332104 (Example 57). Protoplasts are cultured following transformation at a density of 2×10^6 per ml in KM-8p medium with no solidifying agent added.

Measurements of the levels UFGT expression are made using colorimetry and/or biochemical detection methods such as Northern blots or the enzyme activity assays described by Dooner and Nelson, Proc. Natl. Acad. Sci. 74: 5623-5627 (1977). Comparison is made with mock treated protoplasts transformed with a vector only control.

Alternatively, or in addition to, analysing expression of UFGT in transformed protoplasts, intact maize plants may be recovered from transformed protoplasts and the extent of UFGT expression determined. Suitable protocols for growing up maize plants from transformed protoplasts are known in the art: Electroporated protoplasts are resuspended in Km-8p medium containing 1.2% w/v Seaplaque agarose and 1 mg/l 2,4-D. Once the gel has set, protoplasts in agarose are placed in the dark at 26°C. After 14 days, colonies arise from the protoplasts. The agarose containing the colonies is transferred to the surface of a 9 cm diameter petri dish containing 30 ml of N6 medium (EP-A-332,104) containing 2,4-D solidified with 0.24% Gelrite®. 100 mg/l hygromycin B is also added to select for transformed cells. The callus is cultured further in the dark at 26°C and callus pieces subcultured every two weeks onto fresh solid medium. Pieces of callus may be analysed for the presence of the pTMBz23 construct and/or UFGT expression determined.

Corn plants are regenerated as described in Example 47 of EP-A-332,104. Plantlets appear in 4 to 8 weeks. When 2 cm tall, plantlets are transferred to ON6 medium (EP-A-332,104) in GA7 containers and roots form in 2 to 4 weeks. After transfer to peat pots plants soon become established and can then be treated as normal corn plants.

Plantlets and plants can be assayed for UFGT expression as described above.

Example 12 – Regulation of gene expression using a chemically inducible small molecule

The Zif268 Zinc finger phage display library described in Example 1 is screened using the bronze promoter sequence described in Example 11 and a library of small molecule candidate DNA binding ligands, prescreened to remove non-DNA binding molecules. The protocol used is essentially a modification of Example 1 but using multiple ligands. To increase the number of ligands in the screen, ligands are screened in groups of twenty. Once zinc finger clones are identified that have ligand-dependent DNA binding, a single zinc finger clones is tested for ligand-dependent binding against each individual ligand in the mixture originally selected. In this way, a gene switch comprising a zinc finger clone that binds to a region of the bronze promoter in a manner modulatable by a chemical ligand, the region of the bronze promoter and the chemical ligand itself is identified.

The zinc finger clone is fused to a VP16 transactivation domain and other relevant sequences as described in Example 7. The resulting construct, pZFSelectC1 is transferred to the plant binary vector pBIN121 between the Cauliflower Mosaic Virus 35S promoter and the nopaline synthase terminator sequence. The binary construct thus derived is then introduced into *Agrobacterium tumefaciens* (strain LBA 4044 or GV 3101) either by triparental mating or direct transformation.

A transgenic *Arabidopsis* line, designated *At-ZFSelectC1*, is produced as described above in Example 8.

A further transgenic *Arabidopsis* line, designated *At-BzGUS* is produced which comprises a reporter construct containing the *E. coli* beta-glucuronidase gene (GUS) fused to a –90 minimal 35S promoter to which is operably linked the bronze promoter sequence used in the tripartite screen. *Arabidopsis* lacks endogenous GUS activity. Further, GUS activity is very stable and expression can be measured accurately using fluometric assays of very

small amounts of transformed plant tissue (see Jefferson *et al.*, Embo J. 6: 3901-3907 (1987)).

At-ZFSelectC1 lines are crossed with *At*-BzGUS lines. The progeny of such crosses yield plants that carry the reporter construct BzGUS together with either the zinc finger protein construct ZFSelectC1.

Plants are grown in a range of concentrations of the chemical ligand and GUS activity in leaf tissue measured as described in Jefferson *et al.*, Embo J. 6: 3901-3907 (1987). GUS activity in non transgenic plants, *At*-ZFSelectC1 line and *At*-BzGUS lines in the presence of the chemical ligand is also measured.

Example 13 – Tripartite Screen for a zinc finger/target DNA and small molecule ligand and the use of the identified components in regulating gene expression

A screen is performed as described in Example 12 except that the target DNA is a randomised library based on the Bronze promoter sequence and the procedure described in Example 1.3 is used to determine the binding site signature of identified clones once a ligand has been selected. Verification of the target DNA sequence is also performed as described in Example 1.3.

A target DNA identified in the screen is introduced into a -90 minimal Ca35S-GUS reporter construct as described in Example 12 and used to produce a transgenic *Arabidopsis* line. A corresponding zinc finger clone is introduced into an expression construct as described in Example 12 and used to produce a transgenic *Arabidopsis* line. The two lines are crossed and progeny tested for induction of GUS activity in the presence or absence of the ligand identified in the screen.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly

limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Sequence ID 1: TFIIIA/Zif-VP16

TCTAGAGCGCCGCCATGGGAGAGAAGGCGCTGCCGGTGGTGTATAAGCGGTACATCTGCTC
 TTTCGCCGACTGCGGCGCTGCTTATAACAAGAACTGGAAACTGCAGGCGCATCTGTGCAAA
 5 CACACAGGAGAGAAACCATTTCATGTAAAGGAAGAAGGATGTGAGAAAGGCTTTACCTCGC
 TTCATCACTTAACCCGCCACTCACTCACTCATACTGGCGAGAAAACTTCACATGTGACTC
 GGATGGATGTGACTTGAGATTTACTACAAAGGCAAACATGAAGAAGCACTTTAACAGATTC
 CATAACATCAAGATCTGCGTCTATGTGTGCCATTTTGAGAACTGTGGCAAAGCATTCAAGA
 AACACAATCAATTAAAGGTTTCATCAGTTCAGTCACACACAGCAGCTGCCGTATGCTTGCCC
 10 TGTCGAGTCCTGCGATCGCCGCTTTTCTCGCTCGGATGAGCTTACCCGCCATATCCGCATC
 CACACAGGCCAGAAGCCCTTCCAGTGTGCAATCTGCATGCGTAACTTCAGTCGTAGTGACC
 ACCTTACCACCCACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGG
 GAGGAAGTTTGCCAGGAGTGATGAACGCAAGAGGCATACCAAATCCATTTAAGACAGAAG
 GACGCGGCCGCACTCGAGCGGAATTCCGGCCCAAAAAAGAAGAGAAAGGTCGCCCCCCCCGA
 15 CCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACGGCGAGGACGTGGCGATGGCGCATGC
 CGACGCGCTAGACGATTTTCGATCTGGACATGTTGGGGGACGGGGATTCCCCGGGGCCGGGA
 TTTACCCCCACGACTCCGCCCCCTACGGCGCTCTGGATACGGCCGACTTCGAGTTTGAGC
 AGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGTGGGAACAAAACTTATTTCTGA
 AGAAGATCTGTAAGGATCC
 20

Sequence ID 2: TFIIIA/Zif-VP64

TCTAGAGCGCCGCCATGGGAGAGAAGGCGCTGCCGGTGGTGTATAAGCGGTACATCTGCTC
 TTTCGCCGACTGCGGCGCTGCTTATAACAAGAACTGGAAACTGCAGGCGCATCTGTGCAAA
 25 CACACAGGAGAGAAACCATTTCATGTAAAGGAAGAAGGATGTGAGAAAGGCTTTACCTCGC
 TTCATCACTTAACCCGCCACTCACTCACTCATACTGGCGAGAAAACTTCACATGTGACTC
 GGATGGATGTGACTTGAGATTTACTACAAAGGCAAACATGAAGAAGCACTTTAACAGATTC
 CATAACATCAAGATCTGCGTCTATGTGTGCCATTTTGAGAACTGTGGCAAAGCATTCAAGA
 AACACAATCAATTAAAGGTTTCATCAGTTCAGTCACACACAGCAGCTGCCGTATGCTTGCCC
 30 TGTCGAGTCCTGCGATCGCCGCTTTTCTCGCTCGGATGAGCTTACCCGCCATATCCGCATC
 CACACAGGCCAGAAGCCCTTCCAGTGTGCAATCTGCATGCGTAACTTCAGTCGTAGTGACC
 ACCTTACCACCCACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGG
 GAGGAAGTTTGCCAGGAGTGATGAACGCAAGAGGCATACCAAATCCATTTAAGACAGAAG
 GACGCGGCCGCACTCGAGCGGAATTCCGGCCCAAAAAAGAAGAGAAAGGTCGAACCTCAGC
 35 TGACTTCGGATGCATTAGATGACTTTGACTTAGATATGCTAGGATCTGACGCGCTAGACGA
 TTTTCGATCTGGACATGTTGGGCAGCGATGCTCTGGACGATTTTCGATTTAGATATGCTTGGC
 TCGGATGCCCTGGATGACTTCGACCTCGACATGCTGTCAAGTCAGCTGAGCCAGGAACAAA
 AACTTATTTCTGAAGAAGATCTGTAAGGATCC

40 Sequence ID 3: TFIIIA/Zif binding site

TgcgtgggcgTGTACCTggatgggagacC

CLAIMS

1. A method of selecting a gene switch, which gene switch comprises (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand, which method comprises:
 - (a) contacting one or more candidate target DNA molecule(s) with one or more candidate DNA binding molecules, in the presence of one or more DNA binding ligands, wherein at least one of the candidate DNA binding molecules comprises a non-naturally occurring DNA binding domain;
 - (b) selecting a complex comprising a candidate target DNA, a DNA binding molecule and a DNA binding ligand;
 - (c) isolating and/or identifying the unknown components of the complex;
 - (d) comparing the binding of the DNA binding molecule component of the complex to the target DNA component of the complex in the presence and absence of the DNA binding ligand component of the complex; and
 - (e) selecting complexes where said binding differs in the presence and absence of the DNA binding ligand component.
2. A method according to claim 1 wherein the DNA binding molecules are provided as a plurality of DNA binding molecules.
3. A method according to claim 2 wherein the DNA binding molecules are provided as a library of DNA binding molecules.
4. A method according to any one of claims 1 to 3 wherein the target DNA is provided as a plurality of DNA sequences.
5. A method according to any one of claims 1 to 4 wherein the target DNA is provided as a library of DNA sequences, said sequences being related to one another by sequence homology.

6. A method according to any one of the preceding claims wherein a plurality of candidate DNA binding ligands are used.
7. A method according to claim 6 wherein one target DNA sequence is used.
8. A method according to claim 6 or claim 7 wherein one of the components isolated and/or identified in step (c) is a DNA binding ligand component.
9. A method according to any one of the preceding claims wherein one of the components isolated in step (c) is a DNA binding molecule component.
10. A method according to any one of the preceding claims wherein the DNA binding molecule component has a higher affinity for the target DNA in the presence of the DNA binding ligand component than in the absence of the DNA binding ligand component.
11. A method according to any one of claims 1 to 9 wherein the DNA binding molecule component has a higher affinity for the target DNA in the absence of the DNA binding ligand component than in the presence of the DNA binding ligand component.
12. The method according to any one of the preceding claims, wherein said candidate DNA binding molecules are polypeptides.
13. The method according to claim 12, wherein said candidate DNA binding molecules are polypeptides at least partly derived from transcription factors.
14. The method according to claim 13, wherein said candidate DNA binding molecules are derived from zinc finger transcription factors.
15. A method according to any one of the preceding claims, wherein the candidate DNA binding molecules are provided as a phage display library.
16. A method according to any one of the preceding claims, wherein the DNA binding ligand is selected from Distamycin A, Actinomycin D and echinomycin.

17. A gene switch comprising (i) a target DNA molecule; (ii) a DNA binding molecule which binds to the target DNA molecule in a manner modulatable by a DNA binding ligand; and (iii) the DNA binding ligand.
18. Use of a DNA binding molecule selected by the method of any one of claims 1 to 16 in a method of regulating transcription from a DNA sequence comprising a target DNA to which the DNA binding molecule binds in a manner modulatable by a DNA binding ligand.
19. Use of a DNA binding ligand selected by the method of any one of claims 1 to 16 in a method of regulating transcription from a DNA sequence comprising a target DNA to which a DNA binding molecule binds in a manner modulatable by the DNA binding ligand.
20. Use of a target DNA selected by the method of any one of claims 1 to 16 in a method of regulating transcription from a DNA sequence comprising the target DNA to which a DNA binding molecule binds in a manner modulatable by a DNA binding ligand.
21. A method of modulating the expression of one or more genes, said method comprising administering a DNA binding molecule and DNA binding ligand selected according to the method of any one of claims 1 to 16 to a cell wherein the regulatory sequences of said genes comprise a target DNA selected according to the method of any one of claims 1 to 16.
22. A method of modulating the expression of one or more nucleotide sequences of interest in a host cell which host cell comprises a nucleic acid sequence capable of directing the expression of a DNA binding molecule and a target DNA sequence to which the DNA binding molecule binds in a manner modulatable by a DNA binding ligand which method comprises administering said DNA binding ligand to the cell and wherein the DNA binding molecule is heterologous to the host cell.
23. A method according to claim 21 or claim 22 wherein the host cell is a plant cell.

24. A method according to claim 23 wherein the plant cell is part of a plant and the target sequence is part of a regulatory sequence to which the nucleotide sequence of interest is operably linked, said regulatory sequence being preferentially active in the male or female organs of the plant.
25. A non human transgenic organism comprising a target DNA sequence and a nucleic acid sequence capable of directing the expression of a DNA binding molecule which binds to the target DNA in a manner modulatable by a DNA binding ligand wherein the target DNA sequence and/or nucleic acid sequence are heterologous to the organism.
26. A transgenic non-human organism according to claim 25 which is a plant.

Figure 1

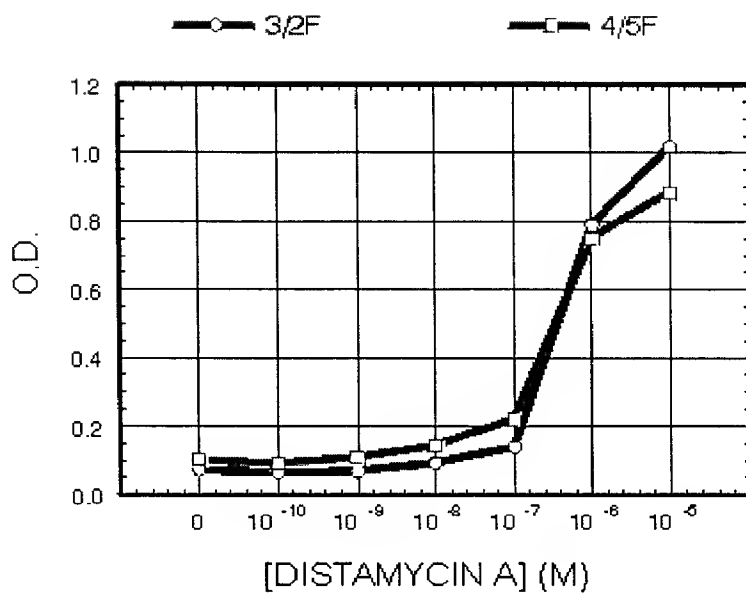


Figure 2

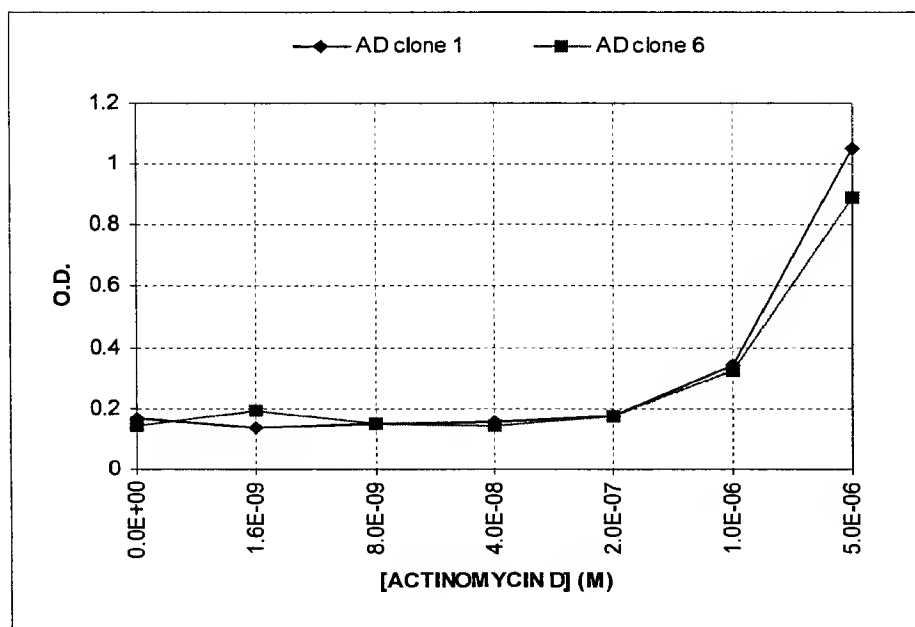


Figure 3

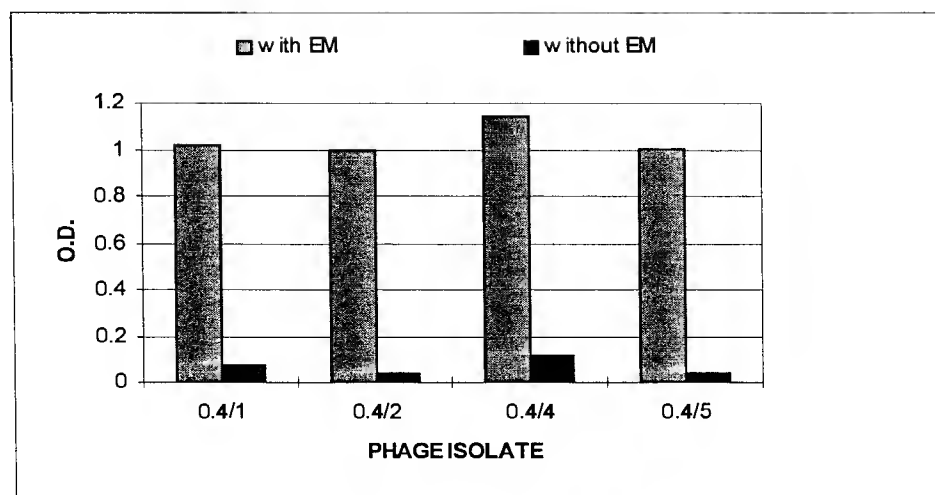


Figure 4

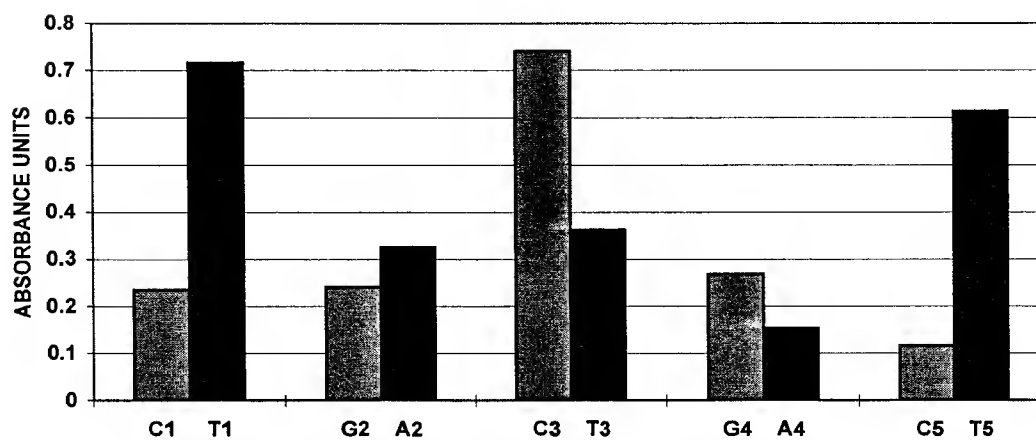


Figure 5

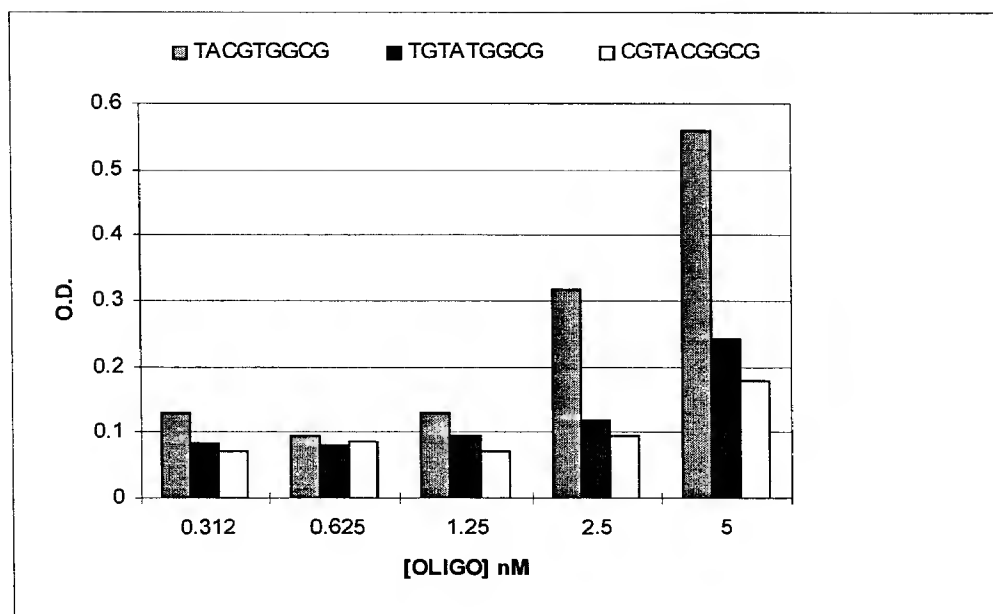
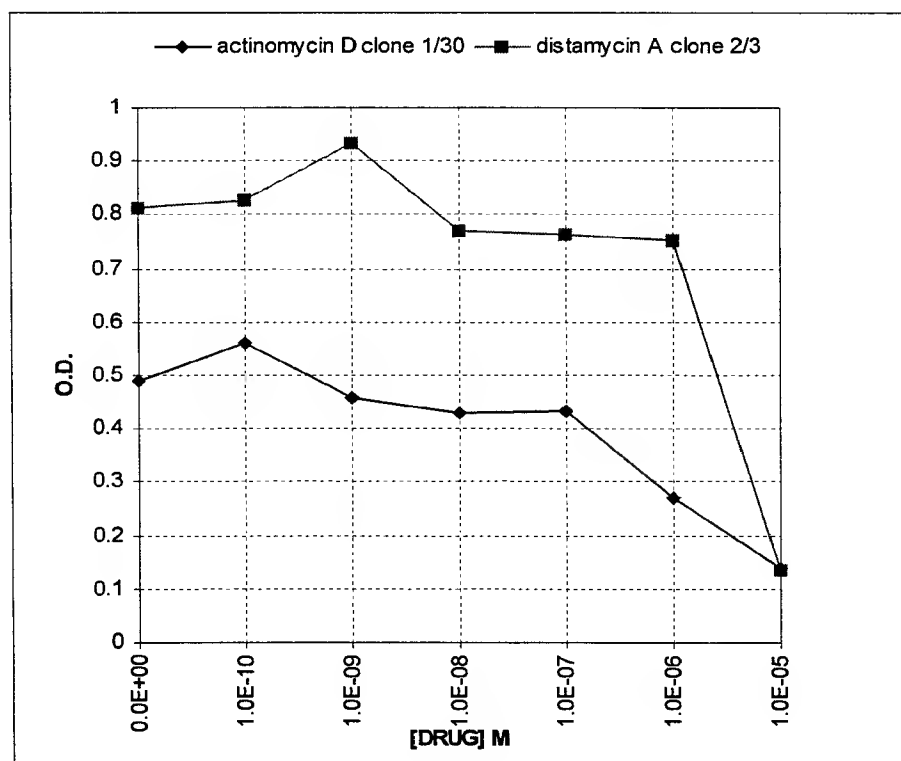


Figure 6



INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/02071

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/10 C12N15/63 C12N15/82 C07K14/47 C12Q1/68
G01N33/50 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, STRAND, EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DICKINSON LILIANE A ET AL: "Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 95, no. 22, 27 October 1998 (1998-10-27), pages 12890-12895, XP002148251 Oct. 27, 1998 ISSN: 0027-8424	17,22
Y	the whole document --- -/-	23-26

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

22 September 2000

Date of mailing of the international search report

10/10/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

PCT/GB 00/02071

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

In ternational Application No

PCT/GB 00/02071

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WO 96 06166 A (MEDICAL RES COUNCIL ;CHOO YEN (SG); KLUG AARON (GB); GARCIA ISIDRO) 29 February 1996 (1996-02-29) cited in the application the whole document -----</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 00/02071

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9429442 A	22-12-1994	AU 684524 B	18-12-1997
		AU 7108194 A	03-01-1995
		CA 2165162 A	22-12-1994
		DE 705334 T	30-12-1999
		EP 0705334 A	10-04-1996
		ES 2140359 T	01-03-2000
		JP 9500526 T	21-01-1997
		US 5650298 A	22-07-1997
		US 5789156 A	04-08-1998
		US 5888981 A	30-03-1999
		US 5859310 A	12-01-1999
		US 6004941 A	21-12-1999
		US 5589362 A	31-12-1996
		US 5814618 A	29-09-1998
		US 5866755 A	02-02-1999
		US 5912411 A	15-06-1999
		US 5922927 A	13-07-1999
WO 9853057 A	26-11-1998	AU 7542298 A	11-12-1998
		AU 7542398 A	11-12-1998
		AU 7542698 A	11-12-1998
		EP 0983349 A	08-03-2000
		EP 0983350 A	08-03-2000
		EP 0983351 A	08-03-2000
		WO 9853058 A	26-11-1998
WO 9606166 A	29-02-1996	WO 9853060 A	26-11-1998
		AU 1003799 A	22-04-1999
		AU 698152 B	22-10-1998
		AU 3229195 A	14-03-1996
		CA 2196419 A	29-02-1996
		EP 0781331 A	02-07-1997
		JP 10504461 T	06-05-1998
		US 6013453 A	11-01-2000
		US 6007988 A	28-12-1999